VOLUME VI JANUARY, 1927 NUMBER 1

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each. The foreign postage is 35 cents per year or 9 cents per copy.

BACK COPIES

The price of back copies is 50 cents each. The two issues of 1922, the four issues of 1923 and those of January, 1924, October, 1925, and January, 1926, are available, to a limited degree, only for the completion of library files. Other back issues are still available for general distribution at fifty cents per copy.

Address all correspondence to the Editor
Information Department
AMERICAN TELEPHONE AND TELEGRAPH COMPANY
195 Broadway, New York, N. Y.

The Bell System Technical Journal

January, 1927

Electromagnetic Theory and the Foundations of Electric Circuit Theory [1]

By JOHN R. CARSON

SYNOPSIS: The familiar equations which are used to solve for the currents and charges in linear networks summarize the inductive analysis of countless observations made upon such networks. Having been arrived at by inductive methods, these familiar equations of Ohm, Faraday and Kirchhoff are substantially independent of the more general electromagnetic theory of Maxwell and Lorentz. The present paper examines the foundation of electric circuit theory from the standpoint of the fundamental equations of electromagnetic theory and a derivation of the former from the latter is made, in the course of which the assumptions, approximations and restrictions tacitly involved in the equations of circuit theory are explicitly stated. The treatment is sufficiently extended as to show how the familiar equation for the simple oscillating circuit and the so-called telegraph equation can be deduced from the Maxwell-Lorentz statement of electromagnetic theory.

ELECTRIC circuit theory, as the term is employed in the present paper, is that branch of electromagnetic theory which deals with electrical oscillations in linear networks; more precisely stated, with the distribution of currents and charges in the free oscillations of the network, or under the action of impressed electromotive forces. The network is a connected set of closed circuits or meshes each of which is regarded as made up of inductances, resistances and condensers, a simplifying assumption which is fundamental to circuit theory.

The great importance of electric circuit theory in electro-technics does not require emphasis; it is not too much to say that it is responsible in no small measure for the rapid development of electrical engineering and is an absolutely essential guide in the complicated technical problems there encountered.

The equations of electric circuit theory in their present form are essentially a generalization of the observations of Ohm, Faraday, Henry, Kirchhoff and others and their development preceded the electromagnetic theory of Maxwell and Lorentz. Naturally, in view of its early development, circuit theory embodies approximations, the precision of which cannot be determined from the observations on which it is based. For example, circuit theory explicitly ignores the finite velocity of propagation of electromagnetic disturbances, and

[1] In its original form this paper was read before the National Academy of Sciences, April 1925. Subsequently it was amplified and revised and included in a lecture course delivered at the Massachusetts Institute of Technology, April 1926.

hence the phenomena of radiation. Again it involves the assumption that the network can be represented by a finite number of coordinates and thus that it constitutes a rigid dynamic system. The rigorous equations of electromagnetic theory formulate the relations between current and charge *densities* and the accompanying fields. Circuit theory, on the other hand, expresses approximate relations between total currents and charges and impressed electromotive forces.

With the rapid development of electro-technics an increasing number of problems is being encountered where the application of classical electric circuit theory is of doubtful validity, or where the conclusions derived from it must be interpreted with great care. Such problems are the result not only of the use of very high frequency in radio-transmission but arise also in connection with the need of a more precise theory of wire transmission.

In view of the foregoing it seems desirable to examine the foundations of circuit theory. This is the problem dealt with in the present paper:—a derivation of the classical circuit theory equations from the standpoint of electromagnetic theory, in the course of which the approximations involved are pointed out.

A second reason, pedagogic in character, is believed to justify the present study. This is, that, as circuit theory is usually taught to technical students no picture of its true relation to electromagnetic theory is given, and the student comes to regard inductance, resistance, capacity, voltage, etc., as fundamental concepts.

To start with our problem in a general form, consider a conducting system of any form whatsoever, in which the *charge density* at any point x, y, z at any time t is denoted by

$$\rho(x, y, z, t) = \rho,$$

and the *vector current density* by

$$\boldsymbol{u}(x, y, z, t) = \boldsymbol{u},$$

the functional notation indicating that the charge and the current density are functions of space and time. At any point in the system let

$$\boldsymbol{E}(x, y, z, t) = \boldsymbol{E}$$

denote the *vector electric intensity*. This we shall suppose to be composed of two parts; thus

$$\boldsymbol{E} = \boldsymbol{E}^\circ + \boldsymbol{E}'. \qquad (1)$$

In this equation $\boldsymbol{E}^\circ$ is the *impressed electric intensity* and $\boldsymbol{E}'$ the *electric intensity due to the reaction of the currents and charges in the system*. Thus $\boldsymbol{E}^\circ$ may be the electric intensity due to a distant system, as in radio transmission, or that due to a generator, battery or other

source of energy. In the following we shall suppose that $\boldsymbol{E}^\circ$ is specified and we shall keep carefully in mind the fact that $\boldsymbol{E}^\circ$ denotes the electric intensity *not* due to the reaction of the system itself. This distinction is extremely important.

We have now to take up the problem of specifying the electric intensity $\boldsymbol{E}'$ in terms of the currents and charges of the system. The necessary relation is furnished by the *Lorentz or retarded potentials*

$$\Phi = \int \frac{\rho(t - r/c)}{r} dv, \qquad \text{(scalar)} \tag{2}$$

$$\boldsymbol{A} = \int \frac{\boldsymbol{u}(t - r/c)}{r} dv, \qquad \text{(vector)}. \tag{3}$$

Interpreting equation (2), Φ is equal to the volume integral of the *retarded* charge density divided by the distance between the point at which Φ is evaluated and the location of the charge. The *retarded* charge density means that at time t we take the value of the charge at the earlier time $t - r/c$, where c is the velocity of light. It is to be understood that ρ and $\boldsymbol{u}$ are the true charge and current density, and displacement currents are not included. Their effect appears in the retardation only. c also is the true velocity of propagation in vacuo. The potential Φ is therefore a generalization of the electrostatic potential into which it degenerates in an unvarying system.

Similarly the *vector potential* $\boldsymbol{A}$ of equation (3) is gotten by a volume integral of the retarded vector current density divided by distance r. As the name indicates it is a vector quantity and in Cartesian coordinates has three components A_x, A_y, A_z.

By means of the equation

$$\boldsymbol{E}' = -\operatorname{grad} \Phi - \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A}, \tag{4}$$

the electric intensity due to the reaction of the system is expressed in terms of the charge and current densities.

Equations (2), (3), (4) and the additional equations

$$\boldsymbol{B}' = \operatorname{curl} \boldsymbol{A}, \tag{5}$$

$$\operatorname{div} \boldsymbol{u} = -\frac{1}{c}\frac{\partial}{\partial t}\rho, \tag{6}$$

(where $\boldsymbol{B}'$ is the magnetic induction due to the currents in the system) are the complete equivalent of Maxwell's equations from which they are immediately derivable.

Aside from the fact that the physical significance of the foregoing equations is deducible by direct inspection, they represent a great step because they are *integral equations* whereas Maxwell's equations are *partial differential equations.* A second advantage is that only *true* currents and charges are involved, the displacement currents of Maxwell being replaced by *retarded action at a distance.* Whatever may be said for or against the physical point of view, this effects a substantial mathematical simplification. The formulation of the fundamental field equations in terms of the retarded potentials is due to Lorentz.

In order to complete the specification of the system we have to formulate the relation between the current density $\boldsymbol{u}$ and the electric intensity $\boldsymbol{E}$. In doing so we shall exclude magnetic matter and shall assume that the conductors obey Ohm's law. This restriction is not necessary but effects a great simplification in both the physical picture and the mathematical formulas.[2] We therefore assume that the conducting system is specified completely by its conductivity

$$g = g(x, y, z),$$

and that

$$\frac{1}{g}\boldsymbol{u} = \boldsymbol{E}. \tag{7}$$

Combining with (1) and (4), we have

$$\frac{1}{g}\boldsymbol{u} = \boldsymbol{E}^{\circ} - \text{grad}\,\Phi - \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A}, \tag{8}$$

which is our fundamental equation.[3] The preceding set of equations, if g and $\boldsymbol{E}^{\circ}$ are everywhere specified, enable us, theoretically at least, to completely solve the problem of the distribution of currents and charges in the system.

Before taking up this problem we shall first derive the energy theorem and then investigate the properties of the field by aid of the retarded potentials.

Starting with equation (8), multiply throughout by $\boldsymbol{u}$, getting

$$\frac{1}{g}\boldsymbol{u}^2 = (\boldsymbol{E}^{\circ}\cdot\boldsymbol{u}) - (\boldsymbol{u}\cdot\text{grad}\,\Phi) - \left(\boldsymbol{u}\cdot\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A}\right),$$

and integrate over the system, getting

$$\int\frac{1}{g}\boldsymbol{u}^2dv = \int(\boldsymbol{E}^{\circ}\cdot\boldsymbol{u})dv - \int(\boldsymbol{u}\cdot\text{grad}\,\Phi)dv - \int\left(\boldsymbol{u}\cdot\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A}\right)dv.$$

[2] See Appendix for the general formulas.

[3] See Appendix for the vector notation employed in this paper.

Remembering that $\boldsymbol{u}$ is expressed in *elm.* units, this becomes

$$\frac{1}{c}D = \frac{1}{c}W - \int(\boldsymbol{u}\cdot\text{grad }\Phi)dv - \int\left(\boldsymbol{u}\cdot\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A}\right)dv$$

or

$$W = D + c\int(\boldsymbol{u}\cdot\text{grad }\Phi)dv + c\int\left(\boldsymbol{u}\cdot\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A}\right)dv,$$

where W is the work done per unit time by the impressed electric field, and D is the ***dissipation*** per unit time in the system; i.e., the rate at which electrical energy is converted into heat. By means of general theorems in vector analysis, the integrals can be transformed and the equation reduced to the form

$$W = D + \frac{\partial}{\partial t}\frac{1}{8\pi}\int(\boldsymbol{E}^2 + \boldsymbol{H}^2)dv + \frac{c}{4\pi}\int[\boldsymbol{E}\cdot\boldsymbol{H}]_n dS,$$

the last integral being taken over any closed surface which includes the system. Translating this equation into words, it states that:—

The work done per unit time by the impressed forces is equal to the rate of dissipation per unit time plus the rate of increase of the field energy plus the rate at which energy is *radiated from* the system. The vector $(c/4\pi)[\boldsymbol{E}\cdot\boldsymbol{H}]$ is the *radiation vector* and gives the density and direction of energy flow per unit time;[4] it will be denoted by $\boldsymbol{S}$.

We now shall briefly consider the field due to the currents and charges in the system.

If the current density $\boldsymbol{u}$ and charge density ρ are everywhere specified, the retarded potentials are uniquely and completely determined by the formulas

$$\boldsymbol{A} = \int\frac{\boldsymbol{u}(t - r/c)}{r}dv, \qquad \text{(vector)}$$

$$\Phi = \int\frac{\rho(t - r/c)}{r}dv. \qquad \text{(scalar)}.$$

The functional notation $\boldsymbol{u}(t - r/c)$ and $\rho(t - r/c)$ indicating that $\boldsymbol{u}$ and ρ are to be evaluated at time $t - r/c$ may profitably be replaced by $\boldsymbol{u}e^{-(p/c)r}$ and $\rho e^{-(p/c)r}$, so that

$$\boldsymbol{A} = \int\frac{\boldsymbol{u}e^{-(p/c)r}}{r}dv,$$

$$\Phi = \int\frac{\rho e^{-(p/c)r}}{r}dv.$$

[4] This is known as Poynting's theorem.

These expressions may be interpreted in either of two ways. (1) If $p = i\omega$ where $\omega = 2\pi f$ and $i = \sqrt{-1}$, then the formulas are the usual complex steady state expressions. On the other hand if p is regarded as d/dt, they are *operational formulas*. It is worth while to explain the latter briefly on account of its own interest and its bearing on the operational calculus.

The differential equations of $\boldsymbol{A}$ and Φ are

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\boldsymbol{A} = 4\pi\boldsymbol{u},$$

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\Phi = 4\pi\rho.$$

where

$$\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2.$$

Now it will be recalled that the differential equation of the electrostatic potential V is

$$\nabla^2 V = 4\pi\rho$$

and that its solution is

$$V = \int \frac{\rho}{r}\,dv.$$

Operationally

$$\left(\nabla^2 - \frac{p^2}{c^2}\right)\Phi = 4\pi\rho$$

and the corresponding solution is

$$\Phi = \int \frac{\rho e^{-(p/c)}}{r}\,dv.$$

Now this is an operational equation in which ρ is an arbitrary time function. Its solution depends on the following general operational theorem.[5]

If x is defined by the operational equation

$$x = f(t)e^{-\lambda p},$$

then

$$x = f(t - \lambda).$$

Consequently, the solution of the operational equation for Φ is[6]

$$\Phi = \int \frac{\rho(t - r/c)}{r}\,dv.$$

[5] See "The Heaviside Operational Calculus," *Bull. Amer. Math. Soc.*, Jan., 1926.

[6] A proof of this theorem by operational methods was privately communicated to the author several years ago by Stuart Ballantine.

Let us now examine the field of the currents and charges by aid of the formulas

$$\boldsymbol{E} = -\operatorname{grad}\Phi - \frac{p}{c}\boldsymbol{A},$$
$$\boldsymbol{H} = \operatorname{curl}\boldsymbol{A}.$$

Performing the indicated operations,

$$\operatorname{curl}\boldsymbol{A} = -\int e^{-(p/c)r}[\boldsymbol{n}\cdot\boldsymbol{u}]\left(\frac{1}{r^2}+\frac{p}{c}\frac{1}{r}\right)dv,$$
$$\operatorname{grad}\Phi = -\int e^{-(p/c)r}\rho\cdot\boldsymbol{n}\left(\frac{1}{r^2}+\frac{p}{c}\frac{1}{r}\right)dv,$$

where $\boldsymbol{n}$ is a unit vector, parallel to $\boldsymbol{r}$, drawn through the contributing element.

We see from these formulas that the magnetic field due to the currents, and the electric field due to the charges, consist each of two components; one varying inversely as the square of the distance from the contributing element and the other inversely as the distance. Writing $p = i\omega = i\cdot 2\pi f$, the orders of magnitude of the two components are $1/r^2$ and ω/c^2 and their ratio is $2\pi(r/\lambda)$, where λ is the wave length.

The first component is the *induction field*, and involves the frequency only through the exponential term; the second is the *radiation* field and involves the frequency linearly.

If we are considering points in the system itself, and if the dimensions of the system are so small that $2\pi(r/\lambda)$ is small compared with unity, the expressions reduce to

$$\operatorname{curl}\boldsymbol{A} = -\int\frac{[\boldsymbol{n}\cdot\boldsymbol{u}]}{r^2}dv,$$
$$\operatorname{grad}\Phi = -\int\boldsymbol{n}\frac{\rho}{r^2}dv.$$

If therefore the dimensions of the system are sufficiently small with respect to the wave length, these expressions can be employed in calculating the distribution of the currents and charges in the system. This is usually the case in circuit theory, even at radio frequencies.

At a great distance from the system, however, the case is quite different. For no matter how large the wave length, λ, if we consider points outside the system such that $2\pi(r/\lambda)$ is everywhere large compared with unity, the second or radiation field will predominate. This leads to the important conclusion that the field which determines the distribution of currents and charges in the system is quite different

from the field which determines the radiation, and explains the fact that radiation may usually be neglected in calculating the distribution in the network.[7]

To examine the radiation field, consider a point P at such a distance from the system that $2\pi(r/\lambda)$ is very large. Choose any point in the system as the origin and write r_0 as the distance from the origin to the point P, and r' the distance from the contributing element P' to the origin. Then

$$r = r_0 - (\boldsymbol{r}'\cdot\boldsymbol{n}),$$

where $\boldsymbol{n}$ is the unit vector parallel to r_0 and

$$\boldsymbol{A} = \frac{e^{-i\omega r_0}}{r_0}\int \boldsymbol{u}\cdot e^{i\omega(\boldsymbol{r}'\cdot\boldsymbol{n})}dv = \frac{e^{-i\omega r_0}}{r_0}\boldsymbol{J},$$

$$\text{curl } \boldsymbol{A} = -\,i\omega\int\frac{e^{-i\omega r_0}}{r_0}[\boldsymbol{n}\cdot\boldsymbol{J}],$$

which determines $\boldsymbol{H}$.

Instead of calculating $\boldsymbol{E}$ from the formula

$$-\,\text{grad } \Phi - \frac{i\omega}{c}\boldsymbol{A},$$

we make use of the fact that in the dielectric

$$i\omega\boldsymbol{E} = \text{curl } \boldsymbol{H},$$

whence

$$\boldsymbol{E} = -\,[\boldsymbol{H}\cdot\boldsymbol{n}].$$

The interpretation of these equations is that in the radiation field $\boldsymbol{E}$ and $\boldsymbol{H}$ are equal, are in phase and are perpendicular to each other and to the vector r_0. Consequently the radiation vector $\boldsymbol{S}$ is given by

$$\boldsymbol{S} = \frac{c}{4\pi}\boldsymbol{H}^2$$
$$= \frac{c\omega^2}{4\pi}\frac{|\boldsymbol{J}|^2}{r^2},$$

and the radiation is everywhere outward.

These formulas can be used to calculate the radiation in terms of the current distribution alone, and the charge distribution does not appear explicitly.

[7] Conversely the field in the immediate neighborhood of the system is no criterion of the radiation field or the radiating properties of the system. This fact is not always kept in mind by radio-engineers.

Derivation of the Familiar Circuit Theory Relations

In the foregoing we have tacitly assumed that the distribution of currents and charges in the systems is known. We now take up the more difficult problem of determining the distribution in terms of the impressed field and the geometry and electrical constants of the system. This will introduce us to circuit theory and the enormous complexity of the general rigorous expressions will show its important role in physics and engineering. In fact without the beautiful simplifications of circuit theory very few problems of this type could be solved.

In taking up this problem there are two possible modes of approach. In accordance with one we start with Maxwell's differential equations and try to find a solution which satisfies the geometry of the system and the boundary conditions. For conducting systems of simple geometrical shapes solutions in this way are possible. For complicated networks, however, this mode of approach is quite hopeless.

The other mode of approach is to start with the equation

$$\begin{aligned} \frac{1}{g}\boldsymbol{u} &= \boldsymbol{E}^\circ - \text{grad}\,\Phi - i\omega\boldsymbol{A} \\ &= \boldsymbol{E}^\circ - \text{grad}\int\frac{\rho(t-r/c)}{r}\,dv - i\omega\int\frac{\boldsymbol{u}(t-r/c)}{r}\,dv, \end{aligned} \tag{8}$$

which, together with the relation

$$i\omega\rho = -\,\text{div}\,\boldsymbol{u},$$

is an integral equation which completely determines the distribution of currents and charges in the system provided g and $\boldsymbol{E}^\circ$ are specified.

For general purposes of calculation it is quite hopeless as it stands. It has, however, several advantages. First, it is a direct and complete statement of the physical relations which obtain everywhere. Second, it uniquely determines the distribution and does not, like the differential equations, involve the determination of integration constants from the boundary conditions. Third, it leads, through appropriate approximations, to the philosophy and equations of circuit theory.

To start with a simple case, the solution of which can be extended without difficulty to the general network, consider a conductor forming a closed circuit. We suppose that it is exposed at every point to an impressed electric force $\boldsymbol{E}^\circ$, and we suppose that the surrounding dielectric is perfectly non-conducting. It is now our problem to derive, for this simple circuit, the circuit equations, in terms of *total currents* and *charges*, from the rigorous integral equation for the current and charge *densities*.

In the interior of the conductor let us assume a curve s defined as parallel, at every point, to the direction of the resultant current. We do not know precisely the path of this curve but we do know that such a curve can be drawn. In the case of wires of uniform cross section it will be approximately parallel to the axis of the wire. Let the cross section of the conductor normal to s be denoted by S. The total current I_s, parallel to s, is then given by

$$I_s = I = \int u_s dS.$$

Now corresponding to the surface S and its element dS, let us define a hypothetical surface Σ and its element $d\sigma$ by the equation

$$u_s dS = I d\sigma,$$

whence

$$\int u_s dS = I = I \int d\sigma = I \cdot \Sigma,$$

so that Σ is always unity. Now multiply the equation

$$\frac{1}{g} u_s = E_s^\circ - \frac{\partial}{\partial s}\Phi - i\omega A_s \tag{9}$$

by $d\sigma$ and integrate over the cross section Σ; we get

$$\int \frac{u_s}{g} d\sigma = \int E d\sigma - i\omega \int A_s d\sigma - \frac{\partial}{\partial s}\int \Phi d\sigma.$$

This can be written as

$$r(s)I(s) = \overline{E}(s) - i\omega \overline{A}_s(s) - \frac{\partial}{\partial s}\overline{\Phi}(s),$$

or simply

$$rI = \overline{E} - i\omega\overline{A} - \frac{\partial}{\partial s}\overline{\Phi}; \tag{10}$$

r is simply the resistance per unit length of the conductor, since

$$rI^2 = \int \frac{u_s^2}{g} dS = \text{dissipation per unit length due to current } I_s,$$

while $\overline{E}$ is the mean impressed electric force, parallel to s, averaged over the surface Σ.

Now consider the term $i\omega\overline{A}$; we have

$$\overline{A} = \int A_s d\sigma = \int d\sigma \int \frac{u'_s}{r} dv, \qquad u' = u(t - r/c)$$

or, neglecting the retardation,

$$\bar{A} = \int d\sigma \int \frac{u_s}{r} dv.$$

We now assume that the "charging" current normal to s is negligibly small in its contribution to the vector potential, whence

$$\bar{A} = \int ds' I(s') \cdot \cos(s, s') \int d\sigma \int d\sigma' \frac{1}{r}$$
$$= \int I(s') \frac{\cos(s, s')}{r} \lambda(s, s') ds',$$

where

$$\lambda(s, s') = \int d\sigma \int \frac{1}{r} d\sigma'.$$

The term $\vec{\Phi} = \int \Phi d\sigma$ of (10) is next to be considered. Writing

$$\rho dS = Q d\tau,$$

where Q is the total charge per unit length, it becomes

$$\int ds' Q(s') \int d\sigma \int \frac{1}{r} d\tau' = \int Q(s') \mu(s, s') ds',$$

and we get finally

$$rI = \bar{E} - i\omega \int I \cdot \cos(s, s') \lambda(s, s') ds' - \frac{\partial}{\partial s} \int Q \cdot \mu(s, s') ds'. \quad (11)$$

This, together with the further relation

$$i\omega Q = -\frac{\partial}{\partial s} I, \quad (12)$$

constitutes an integral equation in the total current $I = I_s$. That is to say, we have succeeded in passing from the rigorous integral equation in the point function *densities* to an approximate integral equation in terms of the total current and charge per unit length of the conductor. The functions λ and μ of this equation, however, while theoretically determinable from the rigorous equation, are not solvable from the approximate integral equation. Indeed they are, strictly speaking, functions of the mode of distribution of the impressed field $E°$. This fact in most cases, however, is of purely academic interest and λ and μ can be approximately evaluated from the geometry of the conductor by assuming a certain distribution of current density over the cross section. With this problem, however, we have no concern here, we are merely concerned to deduce the form of the canonical equations of circuit theory.

Now let us integrate with respect to s, around the closed curve; we get

$$\begin{aligned}\int rIds &= \int \bar{E}ds - i\omega \int Ids \int \cos(s, s')\lambda(s, s')ds' \\ &= V - i\omega \int lIds,\end{aligned} \tag{13}$$

thus defining the *impressed voltage* V, and the inductance per unit length l. Finally, if we assume that this current variation along the conductor is negligibly small, we get

$$I \int rds = V - i\omega I \int lds,$$

which may be written as

$$RI + i\omega LI = V, \tag{14}$$

which is the usual form of the equation of circuit theory for a closed loop.

In deducing (14) from (10) there is one important point which should be noticed. The assumption that the variation in the current I along the conductor is sufficiently small to justify passing from (13) to (14) does not by any means imply that the effect of the distributed charge, which is absent in (14), is negligible. The term $(\partial/\partial s)\Phi$ vanishes in passing from (12) to (13) because the integration is carried around a closed path. Actually comparing the terms $i\omega\bar{A}$ and $(\partial/\partial s)\bar{\Phi}$, we see that their ratio involves the factor $(\omega/c)^2$ which is an exceedingly small quantity even at very high frequencies. Consequently extremely small variations in the current are sufficient to establish charges which can and do profoundly modify the resultant electric field. These, in the case of a closed circuit, are eliminated from explicit consideration by integrating around a closed curve.

This may be illustrated by brief consideration of a second case where the conductor is not closed but is terminated in the plates of a condenser at $s = s_1$ and $s = s_2$ respectively. Making the same assumption as above, after integrating (11) from $s = s_1$ to $s = s_2$, we get

$$RI + i\omega LI + \Phi_2 - \Phi_1 = V, \tag{15}$$

where $\Phi_2 - \Phi_1$ is the difference in Φ between the condenser plates. Assuming these very close together, $\Phi_2 - \Phi_1$ is approximately proportional to the charge on the condenser, that is, to

$$\int Idt = \frac{1}{i\omega} I,$$

and may be written as $I/\omega C$, whence

$$RI + i\omega LI + \frac{1}{i\omega C} I = V, \tag{16}$$

which is the usual circuit equation for series resistance, inductance and capacity.

Extension of the foregoing to networks containing a plurality of circuits or meshes is straightforward and involves no conceptual or physical difficulties, although branch points may be analytically troublesome. These questions will not be taken up, however, as the foregoing is sufficient to show the connection between general electromagnetic theory and circuit theory and to show how circuit equations may be rigorously derived and their limitations explicitly recognized.

The Telegraph Equation

A particularly interesting and instructive application of the preceding is to the problem of transmission along parallel wires and the assumptions underlying the engineering theory of transmission.[8]

Consider two equal and parallel straight wires so related to the impressed field that equal and opposite currents flow in the wires. Here, corresponding to equation (11), we have

$$\begin{aligned} rI = \bar{E} - i\omega \int I\{\lambda(s, s') - \lambda'(s, s')\} ds' \\ - \frac{\partial}{\partial s} \int Q\{\mu(s, s') - \mu'(s, s')\} ds'. \end{aligned} \tag{17}$$

In this equation $\lambda(s, s')$ is the "mutual inductance" between points s, s' in the same wire while $\lambda'(s, s')$ is the corresponding mutual inductance between point s in one wire and point s' in the other. μ and μ' have a corresponding significance as "mutual potential coefficients."

Now $\lambda(s, s') - \lambda'(s, s')$ is a rapidly decreasing monotonic function of $|s - s'|$ and the same statement holds for $\mu - \mu'$. In view of this property and further assuming the variation of I and Q with respect to s as small, (17) to a first approximation may be replaced by

$$rI = \bar{E} - i\omega I \int (\lambda - \lambda') ds' - \frac{\partial}{\partial s} Q \int (\mu - \mu') ds'. \tag{18}$$

At a sufficient distance from the physical terminals of the wires the

[8] For an entirely different treatment of this problem, reference may be made to "The Guided and Radiated Energy in Wire Transmission," *Trans. A. I. E. E.*, 1924.

integrals become independent of s and approach the limits

$$\int_{-\infty}^{\infty} (\lambda - \lambda')ds' = l,$$

$$\int_{-\infty}^{\infty} (\mu - \mu')ds' = \frac{1}{c},$$

whence

$$rI + i\omega lI + \frac{1}{i\omega c} I = \bar{E}.$$

Finally assuming that the impressed electric intensity $\bar{E} = 0$, and introducing the relation

$$i\omega Q = -\frac{\partial}{\partial s} I,$$

we get

$$\left(r + i\omega l + \frac{1}{i\omega c}\frac{\partial^2}{\partial s^2} \right) I = 0,$$

which is the *telegraph equation*.

Besides its formal theoretical interest the foregoing derivation of the telegraph equation admits of some deductions of practical importance. These deductions, which are rather obvious consequences of the analysis, may be listed as follows.

1. The telegraph equation, as derived above, applies with accuracy only at points at some distance from the physical terminals of the line.
2. The accuracy of the telegraph equation in formulating the physical phenomena decreases in general with increasing frequency.
3. The telegraph equation is the first approximate solution of an integral equation. The first approximate solution decreases in accuracy with decreasing wave length of the propagated current.
4. *While the telegraph equation indicates a finite velocity of propagation of the current along the line, it is based on the assumption that the fields of the currents and charges* (as derived from the potential functions Φ and $\boldsymbol{A}$) *are propagated with infinite velocity.*
5. As a consequence of (4), the telegraph equation does not take into account the phenomena of *radiation*, and in fact indicates implicitly the absence of radiation.

The Coil Antenna

An important example of the type of problem to which the foregoing analysis is applicable is the coil antenna. To this problem

equations (11) and (12) immediately apply but, at least at high frequencies, the approximations introduced above to derive the telegraph equation are not legitimate. This is due to the geometry of the conductor, and also to the fact that the impressed field is not approximately concentrated but is distributed over the entire length of the coil. It is intended to apply these equations to a detailed study of this problem. In the meantime, however, it may be noted that the current depends *not only on the line integral of the impressed electric intensity but also on its mode of distribution along the length of the coil.* This fact may possibly have practical significance in the design of coil antenna and their calibration at very short wave lengths.

Appendix

In the beginning of this paper, it was stated that the analysis applied only to the case of conductors of unit permeability and specific inductive capacity which obey Ohm's Law. The reason for this restriction and the formal extension of the analysis to the more general case will now be briefly discussed.[9]

Suppose that the conductor, instead of having the restricted properties noted above, obeys Ohm's Law but has a permeability μ and specific inductive capacity k which may differ from unity.

The equation (1),

$$\boldsymbol{E} = \boldsymbol{E}^\circ - \text{grad}\,\Phi - i\omega \boldsymbol{A}, \tag{1}$$

still holds, as do also the potential formulas (2) and (3) and the formulas for the electric and magnetic intensities (4) and (5). The relation

$$-\,i\omega\rho = \text{div}\,\boldsymbol{u}$$

is also valid.

The equation $\boldsymbol{u} = g\boldsymbol{E}$ must, however, be modified in the following manner. If we write

$$\boldsymbol{P} = \frac{k-1}{4\pi}\boldsymbol{E},$$

$$\boldsymbol{M} = \frac{\mu - 1}{4\pi\mu}\boldsymbol{H},$$

then the foregoing equations are correct, provided we substitute for the equation $\boldsymbol{u} = g\boldsymbol{E}$ the more general expression

$$\boldsymbol{u} = g\boldsymbol{E} + i\omega\boldsymbol{P} + \text{curl}\,\boldsymbol{M}.$$

[9] For a previous discussion, see "A Generalization of the Reciprocal Theorem," *B. S. T. J.*, July, 1924.

By aid of these relations, the problem involves the solution of the simultaneous integral equations

$$\boldsymbol{E} = \boldsymbol{E}^\circ - \text{grad}\,\Phi - iw\boldsymbol{A},$$

$$\boldsymbol{H} = \boldsymbol{H}^\circ + \text{curl}\,\boldsymbol{A}.$$

These simultaneous equations can immediately be reduced to a single integral equation in $\boldsymbol{u}$, the formal solution of which is straightforward. A study of this equation, however, has not been carried far enough to justify further discussion in the present paper.

Note on Vector Analysis and Notations

In the foregoing, vectors are indicated by Clarendon, or bold-faced type. To those unfamiliar with vector analysis the following may be helpful:

grad Φ is a vector with the Cartesian components

$$\text{grad}_x\,\Phi = \frac{\partial}{\partial x}\Phi, \qquad \text{grad}_y\,\Phi = \frac{\partial}{\partial y}\Phi, \qquad \text{grad}_z\,\Phi = \frac{\partial}{\partial z}\Phi;$$

curl $\boldsymbol{A}$ is a vector with the Cartesian components

$$\text{curl}_x\,\boldsymbol{A} = \frac{\partial}{\partial y}A_z - \frac{\partial}{\partial z}A_y,$$

$$\text{curl}_y\,\boldsymbol{A} = \frac{\partial}{\partial z}A_x - \frac{\partial}{\partial x}A_z,$$

$$\text{curl}_z\,\boldsymbol{A} = \frac{\partial}{\partial x}A_y - \frac{\partial}{\partial y}A_x;$$

div $\boldsymbol{u}$ is a scalar; in Cartesian notation

$$\text{div}\,\boldsymbol{u} = \frac{\partial}{\partial x}u_x + \frac{\partial}{\partial y}u_y + \frac{\partial}{\partial z}u_z.$$

$(\boldsymbol{E}\cdot\boldsymbol{u})$ denotes the *scalar product* of the vectors $\boldsymbol{E}$ and $\boldsymbol{u}$ and itself is a scalar. In Cartesian notation

$$(\boldsymbol{E}\cdot\boldsymbol{u}) = E_xu_x + E_yu_y + E_zu_z.$$

$[\boldsymbol{E}\cdot\boldsymbol{H}]$ denotes the *vector product* of the vectors $\boldsymbol{E}$ and $\boldsymbol{H}$. It is

itself a vector with the Cartesian components

$$
\begin{aligned}
[\boldsymbol{E}\cdot\boldsymbol{H}]_x &= E_yH_z - E_zH_y,\\
[\boldsymbol{E}\cdot\boldsymbol{H}]_y &= E_zH_x - E_xH_z,\\
[\boldsymbol{E}\cdot\boldsymbol{H}]_z &= E_xH_y - E_yH_x.
\end{aligned}
$$

The symbol ∇^2 denotes, in Cartesian coordinates, the operator

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Toll Switchboard No. 3

By J. DAVIDSON

IN the early days of telephony the toll signaling apparatus consisted of a magnetic drop in the line and a drop or ringer in the cord. With the advent of common battery signaling in the local plant, relays and lamps replaced the old type drops and the subscriber was given means for calling the toll operator on a toll connection by operating the switchhook instead of ringing. Up to this time the toll operators were located at the local switchboard and had direct access to the subscriber's line, but with the growth of toll and local traffic, it was no longer economical to place the toll operators at the local board. This led to the development of a separate toll switchboard called the No. 1 board, which had access to the subscriber's line over switching trunks between the toll and local boards. For many years the No. 1 switchboard filled the needs of the time but with the expansion of the toll service and the growth of machine switching local service, it became evident that new arrangements were desirable. The No. 3 toll switchboard was developed to meet the new requirements and it has the following advantages as new installations are required.

(*a*) Reduction in apparatus, resulting in equipment economies.
(*b*) Improved maintenance arrangements.
(*c*) More readily adapted to modifications required by new operating methods.

In discussing the features of the No. 3 board, frequent comparisons will be made with the No. 1 switchboard to set forth the changes which have been made in the design of the new circuits.

Main Features

Cord Simplified by Locating Supervisory Relays in Line and Trunk Circuits

The cord circuit of the No. 1 switchboard is equipped with two supervisory relays. One of these relays responds to 20-cycle current and gives the toll operator a ringing signal, indicating that the distant operator is calling. The second relay responds to direct current received from the switching trunk and gives the operator switchhook supervision of the subscriber. Associated with these two relays are other relays which prevent false signals, and permit the operator to

make a busy test or use the cord for a terminating or a through connection. This cord is shown in Fig. 1.

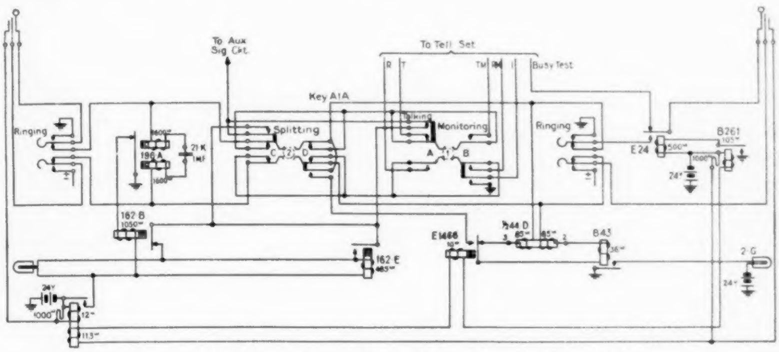


Fig. 1—High impedance toll cord for toll switchboard No. 1

In the No. 3 switchboard the ringing relay and the direct-current supervisory relay, which were formerly connected across the tip and ring conductors of the cord circuit, have been moved from the cord to the line and switching trunk, respectively, and the cord circuit has been simplified as is illustrated in Fig. 2. In this board the line and trunk

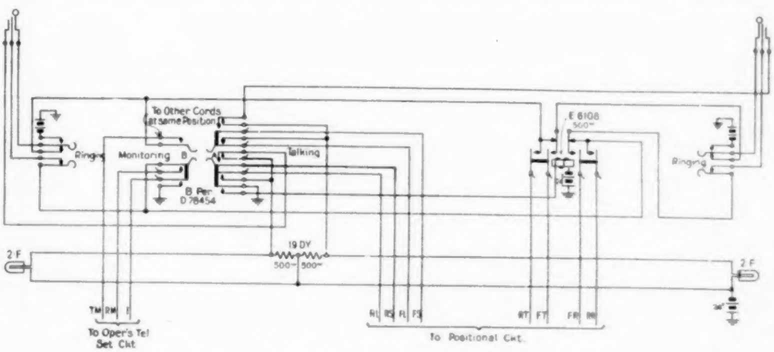


Fig. 2—Toll cord for toll switchboard No. 3

signals are transferred to the cord over the sleeve circuits. This is accomplished by using a nominal sleeve resistance of 1,800 ohms for the line and trunk circuits and connecting the lamps in the sleeves of the cord. Under these conditions there is not sufficient current flowing in the sleeve to light the lamp, but when a ringing signal is received over a line and a cord is associated with that line or when a receiver-on-the-

hook signal is received over a switching trunk, the sleeve resistance of the line or trunk is changed from 1,800 ohms to 80 ohms, which increases the current in the sleeve of the cord sufficiently to light the lamp.

Line Relay Functions in Twofold Capacity

The majority of toll lines in the plant today are of the ringdown type and the operator at one end calls the operator at the distant end by ringing over the line. To receive this ringing signal in the No. 1 board, the lines are equipped with relays which respond to the ringing current received from the distant end of the line and give a line signal. After the operator answers this signal by connecting a cord to the line, the line relay is disconnected and replaced by the ringing relay in the cord which responds to further rering signals over the line. This arrangement of the line and cord, as well as the switching trunk for the No. 1 board, is shown schematically in Fig. 3.

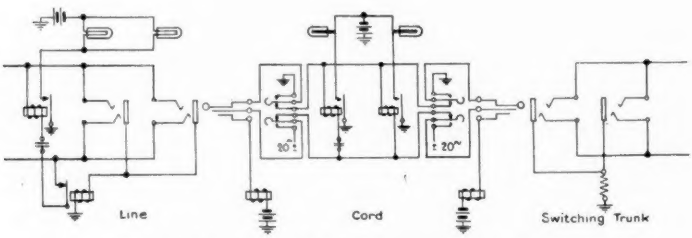


Fig. 3—Schematic: Toll switchboard No. 1 circuits

By transferring the ringing relay from the cord to the line in the No. 3 toll board, this relay is made to function in a twofold capacity, that is, to give the line signal as well as the cord rering signal. When a call is received from a distant point, the apparatus in the line functions to light the line signal and this remains lighted until a toll cord is inserted in the line jack. Further signals over the line cause the apparatus in the line to light the lamp in the cord. This is obtained by changing the sleeve resistance of the line from 1,800 ohms to 80 ohms and is illustrated in schematic form in Fig. 4. As in the past, the line signal is multipled before several operators and appears as a steady illuminated lamp which is extinguished by an operator answering the call. The cord signal appears before one operator and has been changed from a steady lamp signal to a flashing signal for the purpose of obtaining prompt attention on the part of the operator. The cord signal is extinguished when the operator connects to the circuit by the

operation of the talking key. This connects an additional 600 ohms in the sleeve circuit, which releases relays which are held operated in the line circuit and control the lamp.

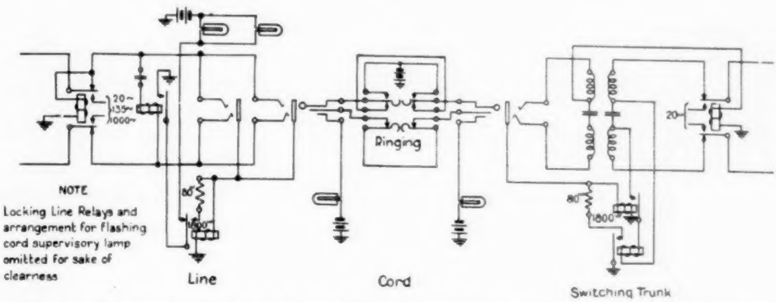


Fig. 4—Schematic: Toll switchboard No. 3 circuits; monitoring and positional circuit keys are not shown

Composite Ringer Simplified

In order that the toll lines may be used for telegraph as well as telephone service, composite sets are often connected into the line circuit at each end. These composite sets are electrical filters which separate the telephone and telegraph currents and direct the telephone currents to the switchboard and the telegraph currents to the telegraph equipment.

When composite sets are connected in the lines terminating in a No. 1 switchboard, it is also necessary to connect a composite ringer in the circuit between the composite set and the switchboard. This is necessary because the 20-cycle current, which is used as ringing current from the switchboard, is in the telegraph range of frequencies and consequently will not pass through the telephone branch of the composite set. The composite ringer substitutes for the 20-cycle outward ringing current received from the switchboard, a higher frequency current which will pass through the telephone path of the composite set. Likewise on incoming ringing signals, the ringer substitutes for the higher frequency current which comes over the line and through the telephone path of the composite set, a 20-cycle current which will operate the ringing relays of the line or cord circuits. A schematic of the composite set and composite ringer, as used with the No. 1 board, is shown in Fig. 5.

In general, the composite ringer for the No. 3 switchboard has been greatly simplified and made a part of the terminating line equipment. This has been accomplished, as illustrated schematically in Fig. 4, by arranging the line circuit so that a relay may be cross-connected in the

line to receive the 20-cycle, or the higher frequency ringing current, and arranging this relay so that it gives the line signal or the cord supervisory signal direct without going through the step of changing ringing frequencies.

Furthermore, the practice of using 20-cycle current in the cord circuit for ringing has been discontinued and ringing is effected in the No. 3 switchboard by connecting 24-volt direct current through the

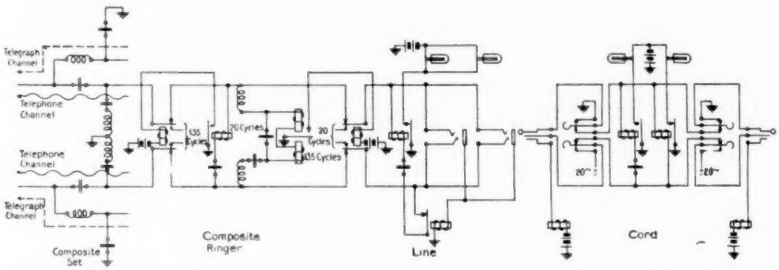


Fig. 5—Schematic: Composite ringer and composited toll line for toll switchboard No. 1

ringing key to the tip conductor of the cord. This current operates a relay of the line or trunk circuit which applies the proper frequency of ringing current to the line or trunk circuit. By this arrangement one relay in the line circuit accomplishes the same result as was accomplished by several relays in the composite ringer. As the ringing current leads to the relay in the line are brought through terminals on the frames, the line can be readily changed for any desired frequency of ringing current.

Elimination of Transfer Key from Face of Inward Switchboard

In the past the practice has been to provide one or two transfer keys per line for each multiple appearance of the line lamp at the inward toll switchboard. The function of these keys is to transfer the inward call from the inward switchboard to the outward delayed positions or to the through positions. With the No. 3 toll switchboard, the use of these transfer keys individual to the line and appearing in the face of each section of the inward switchboard has been discontinued and the transfer is effected by a transfer key in the positional circuit which may be used to transfer a call on any line. This key applies 24-volt battery either directly or through a resistance to the ring conductor of the line and operates the proper transfer relay in the toll line and causes lamps individual to that line to light at the out-

ward, or through positions. This feature not only effects a saving in equipment but saves the space in the face of the switchboard which was formerly occupied by the transfer keys.

Use of Positional Circuit

Another circuit feature of the No. 3 switchboard which marks an improvement over switchboard No. 1 is the use of a so-called positional circuit in which is located much equipment such as splitting keys, dialing keys, etc., which heretofore were individual to each cord. Under normal conditions the tip and ring conductors of the front cord are connected to the tip and ring conductors of the corresponding back cord with no shunts across the circuit. This is illustrated in Fig. 6. By the operation of the talking key associated with each cord

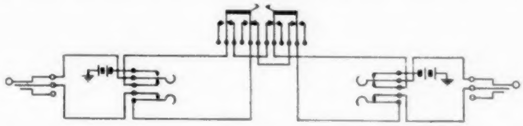


Fig. 6—Schematic: Toll cord talking circuit; talking key normal for toll switchboard No. 3

circuit, the positional circuit is connected between the front and back cords and the operator's telephone set is connected across the circuit as illustrated in Fig. 7. With the talking key of any cord operated, the operator may

(*a*) Dial on either the front or the back cord.
(*b*) Split the talking circuit between the front and the back cords.
(*c*) Transfer an inward call from the inward to the outward or the through positions.

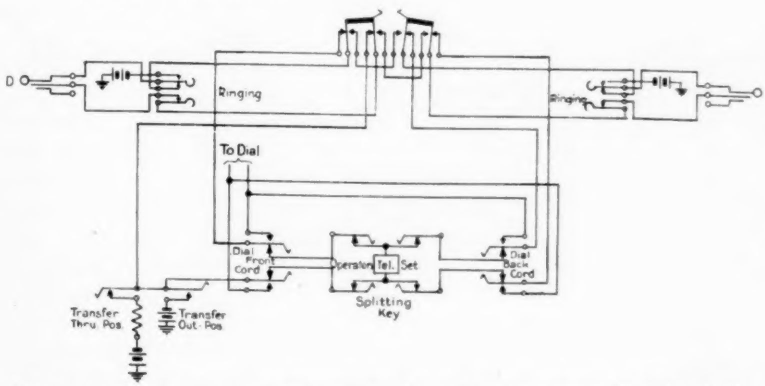


Fig. 7—Schematic: Toll cord positional circuit; talking key operated for toll switchboard No. 3

This circuit arrangement not only effects substantial economies but it is much more flexible and will lend itself to new developments without requiring changes in the cord circuit.

Monitoring and Ringing Keys Individual to Cords

The monitoring and ringing keys are, as in the past, individual to each cord.

Switching Trunk Features

In the No. 3 toll switchboard a repeating coil which has a high impedance to 20-cycle ringing current is used in the outgoing end of the switching trunk. This arrangement has equipment and signaling advantages. Also where loaded toll switching trunks are involved, the use of a repeating coil of the type referred to, but having the proper transmission characteristics, has the advantage of reducing reflection losses by providing for a uniform terminal impedance of the switching trunks.

Principal Advantages

Equipment Economies

As has been pointed out, the expansion in toll business, together with recent developments in the telephone art, have been such that with the circuit arrangements used in the past there has been a growing necessity to add equipment to the cord circuit with the result that the positions are becoming congested with apparatus. With the circuit arrangements outlined for the No. 3 toll switchboard, however, the transfer of the signaling apparatus from the cord to the line and switching trunk makes a marked simplification in the cord and incidentally reduces the congestion in the section. Also it should effect a substantial economy in equipment because of the fact that we are approaching a situation where there are approximately 60 per cent. more cords than lines and 25 per cent. more cords than switching trunks.

The use of the positional circuit and the elimination of the individual splitting key from the toll cord has simplified the switchboard keyshelf. This simplification together with the equipment savings effected by the simplification of composite ringers and the transfer of the supervisory relay equipment from the toll cord to the toll line and switching trunk circuits has effected substantial economies.

Maintenance

In addition to the saving in first cost of equipment the No. 3 switchboard facilitates maintenance. The ordinary toll cords in an office

must be suitable to work with any toll line terminating at the switchboard and consequently with the circuit arrangement used in toll switchboard No. 1, the ringing relay in all the toll cord circuits must be maintained to operate in connection with the longest as well as the shortest line circuit. In the case of the No. 3 toll switchboard, however, the ringing relay is individual to the line and consequently may be adjusted to meet the operating conditions of that line. Long lines with severe ringing conditions require the relay to have a sensitive adjustment while short lines with easy ringing conditions permit a less sensitive relay adjustment to be used which is more easily maintained.

Easily Adaptable to Machine Switching Methods

The introduction of machine switching requires provision for dialing on the trunks and may in the future require the same feature for dialing over toll lines. Such provision in the boards previously employed requires the addition of the necessary keys and relays on a "per cord" basis, whereas with the No. 3 board the equipment can be placed in the positional circuit, without any change in the cord circuit. This results in a great economy in apparatus and makes a change to a dialing basis rather simple.

Summary

It is interesting to note in conclusion that heretofore an increase in cord circuit apparatus has necessarily followed the development of new and improved switchboard systems and the extension of the area of long distance communication. For example, the magneto cord with a single drop bridged across the circuit sufficed in the early days of small magneto boards. The advent of the common battery multiple switchboard brought the necessity for extending switchhook supervision to the toll operator, and resulted in the condenser-type cord consisting of 5 relays, now largely abandoned because of the relatively large transmission loss introduced by it. The high-efficiency cord consisting of 8 relays resulted from the demand for a cord having a minimum transmission loss, and additional complications have resulted in the requirement for dialing in machine switching areas, each improvement, of course, increasing the number of relays in the cord circuit. The No. 3 system, on the other hand, makes possible by the transfer of apparatus to the line and switching trunk and by the use of common positional equipment the relatively simple toll cord shown in Fig. 2 in which the individual apparatus is limited to two keys and one

relay per cord. This provides in many cases a toll cord suitable for either inward, outward or through operation, reduces the apparatus congestion in the section and results in decreased maintenance, while being easily adapted to the future trend in toll development.

The Location of Opens in Toll Telephone Cables

By P. G. EDWARDS and H. W. HERRINGTON

SYNOPSIS: Improved methods have recently been developed for the location of opens in toll cable conductors. The discussion of these methods is prefaced by a review of older practices.

This improved open location method and equipment are sufficiently accurate that in practically all cases a fault in a 60-mile length of cable may be located within a maximum variation of plus or minus one half the length of a cable section (a section is the length of cable between splices—about 750 feet), and therefore enables one to select, prior to the opening of the cable, one or the other of the two splices between which the fault lies. This degree of accuracy is very desirable for practical reasons.

In this development, the line characteristics are considered. The accuracies of calculated locations, assuming no errors in measurements, are compared for different lengths of lines. The impedance bridge circuit is treated to bring out the method of obtaining a balance. The effects of several frequencies of testing potential are analyzed. The probable errors and inaccuracies of measurement which would interfere with the correct location of faults are classified and methods for their correction are developed. The accuracy of the method and the sensitivity of the apparatus are given.

THE location of "opens," or breaks in the continuity of telephone conductors, has always been an important problem in the testing and maintenance of the toll cables of the telephone plant. Although the number of opens encountered in the cable circuits is relatively small as compared to aerial wire lines, their location is more difficult because of certain electrical characteristics of the cable circuits. This condition, coupled with the fact that any work on a toll cable may interfere with a large number of facilities, renders the quick and accurate location of such faults imperative.

A high resistance voltmeter was used in the first feasible attempt to locate opens by means of electrical measurements. The original method of locating opens by the use of a voltmeter consisted essentially of a series of continuity tests. The faulty conductor was connected to ground or to the other wire of the pair at succeeding test stations until the fault was isolated between two adjacent test stations. The trouble was then found, either by inspection of the entire line between test stations, or by further continuity tests with a lineman, first near the middle of the section and later at points gradually approaching the location of the fault.

The inconvenience of such a procedure, however, led to an improved use of the voltmeter. The method employed afforded a rough comparison of the capacitance to ground of the portion of the faulty wire adjacent to the measuring station with that of its good mate or another conductor (of like gauge) following the same route. This was done by allowing the wire to become charged through a volt-

meter and grounded battery, and noting the amount of momentary deflection on the good and bad conductors, respectively, when the polarity of the battery was reversed. The ratio of these deflections gave a general indication of the location of the fault. This method, while giving more accurate results than the continuity test, was still only an approximation and as such was materially affected by line conditions. An appreciable error was produced by leakage due to trees and other causes, and much depended on the judgment of the tester.

Later, the Wheatstone bridge [2] largely displaced the voltmeter. A standard capacitance was compared with the impedance between the open conductor and ground, by varying the ratio arms of the impedance bridge. In this comparison it was necessary to employ some form of alternating testing potential. At first, the simple expedient prevailed of reversing the bridge battery, as in the voltmeter test, the bridge being balanced until no transient unbalance current, or "kick," was indicated by the galvanometer when the battery was reversed. However, when the battery was reversed rapidly, the galvanometer displayed a tendency to stand still at all times. A considerable improvement in the method was effected by providing or the reversal of the galvanometer connections at the same time the battery was reversed. With this arrangement, the galvanometer always read in the same direction, and a balance could be more easily obtained.

Later a relay system was arranged for automatically reversing the connections to the galvanometer as well as for reversing the testing battery. This arrangement relieved the operator of the necessity of doing the reversing manually. Following this, a source of 20-cycle ringing voltage was used for the bridge and also for operating the galvanometer reversing relay. With this arrangement open locations, made on aerial wire lines and short lengths of cable, were fairly satisfactory. However, with the extensive installation of the long toll telephone cables, it was found that open locations made on this type of conductors, did not give a consistently accurate indication of the location of the fault.

As a preliminary step in the development of a suitable open location method and associated apparatus, an analysis was made of all errors which, in general, might enter into a open location. These errors can be classed in several groups for treatment or correction. One group

[2] References: Frank A. Laws, "Electrical Measurements," 1917, McGraw-Hill Book Co., Inc., N. Y.; page 381, "Bridge Measurements of Capacity and Inductance."

"Bridge Methods for Alternating-Current Measurements," D. I. Cone, *Transactions of A. I. E. E.*, July, 1920.

includes errors which are small in comparison to the accuracy of the testing equipment. Errors placed in this group obviously require no compensation. Another group of errors is produced by, or is characteristic of, certain designs of testing equipment. This class of errors has been reduced to negligible magnitude by a redesign of the testing equipment. One general group of errors results from faulty manipulation of the testing equipment or mistakes of computation. This group has been minimized by a convenient arrangement of testing equipment and by outline forms for use in computation.

Another class of errors is introduced by irregularities in the lines or cables on which open locations are made. Some of these are capable of compensation by constant correction factors included in formulae used for computation. Other errors of this class are found to be irregular functions of the length of line, and for their correction or compensation curves have been prepared for each type or condition of irregularity which can be used in the computation of the open location. To simplify the application of the corrections, the curves are so drawn that the correction is given as a simple multiplier.

The preparation of other types of corrections will be developed later in connection with the analysis and treatment of certain specific errors.

In the development of a more sensitive and reliable method of locating opens in telephone lines and cables, it was necessary to make an exact study of the electrical constants of the several types of conductors on which open locations are required. This involved the capacitance and leakance of the conductor to ground, or to neighboring conductors, as well as the series resistance and inductance. The general formula for the impedance of a line open at the distant end is

$$Z_l = Z_0 \coth \theta, \tag{1}$$

where Z_0 is the characteristic impedance of the line and $\theta = Pl$, where P is the propagation constant and l is the length. More fully

$$P = \sqrt{(R + j\omega L)(G + j\omega C)},$$

where R is the series resistance, L the inductance, G the leakance, and C the capacitance, all expressed in terms of the same unit length, and $\omega = 2\pi f$.

In formula (1)

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}},$$

the terms having the same significance as above.

In cable circuits, G is usually zero, and for non-loaded lines, L is

also practically zero. For loaded lines, the inductance is effectually that of the loading coils, and for the frequencies usually employed in open location tests it can be considered as being uniformly distributed. For non-loaded lines, the equation of line impedance is reduced to one of series resistance and capacitance to ground. These constants can be determined by the measurement of short lengths of cable. In making capacitance measurements, the remaining three wires of the quad should be grounded to eliminate their capacitances to ground. When the other three wires of the quad are left free, the capacitance to ground of the faulty wire increases as the length of good wire beyond the break increases. This effect is shown in Fig. 1.

The values of R and C obtained by measuring short lengths of cable are used to calculate P and Z_0. Where the lines are loaded, the nominal inductance and resistance of the loading coils are used and the characteristic constants R and C are, if possible, determined from non-loaded conductors. These constants for one particular cable are listed as follows:

TABLE 1

Constant per Mile	Grade of Loading			
	Non-Loaded	H-44-25	H-174-106	H-245-155
	19-Gauge Inner Layer Conductors			
R	42.90	44.65	47.78	50.34
L	.000+	.015	.062	.089
C	.100	.100	.100	.100
	19-Gauge Outer Layer Conductors			
R	44.00	45.75	48.88	51.44
L	.000+	.015	.062	.089
C	.110	.110	.110	.110
	16-Gauge Conductors			
R	21.00	22.75	25.88	28.44
L	.000+	.015	.062	.089
C	.100	.100	.100	.100

These constants represent values per single-wire mile, R being in ohms, L in henries and C in microfarads.

In making an open location, two impedance measurements are made. One measurement is made on the faulty wire. The other measurement is made on a good wire which follows the same route as the faulty wire. The input impedance of the open conductor divided by the input impedance of the good wire gives an indication of the location of the fault. For short cables, the impedance measured to ground may be regarded as identical with the capacitance com-

ponent of the impedance, because the resistance component of the impedance is negligibly small. However, as the length of cable increases, the input impedance can no longer be regarded as equivalent to the capacitance component of the impedance between the conductor

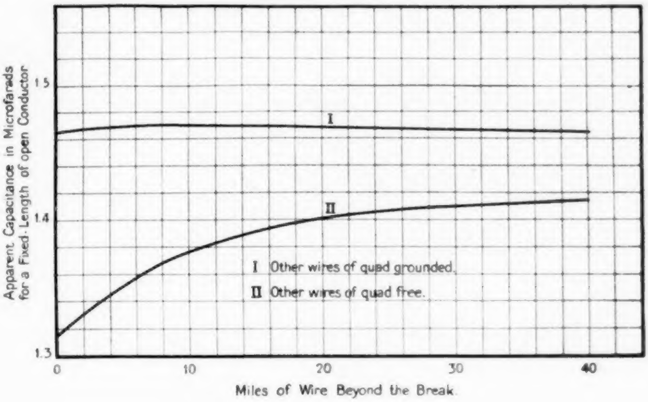


Fig. 1

and ground, because the input impedance is not proportional to the capacitance but varies as the hyperbolic cotangent (Formula 1). The significance and magnitudes of errors inherent in this relation are analyzed and discussed below.

In a homogeneous line, then, if Z_1 and Z_2 represent the input impedances of the bad and good wires respectively, the percentage distance to the fault is most accurately determined by separating the impedances into their real and imaginary parts.

Thus

$$Z_1 = a_1 - jb_1$$

and

$$Z_2 = a_2 - jb_2.$$

If the corresponding lengths are respectively l_1 and l_2, the true distance ratio is l_1/l_2. Assuming that b_2/b_1 is the ratio given by measurement (the shorter line having the higher capacitive reactance),

$$\frac{l_1}{l_2} - \frac{b_2}{b_1} = c,$$

where c is a correction which must be added to or subtracted from the ratio of capacitive reactances to secure the ratio of total capacitance, i.e., the distance ratio.

It is necessary, then, to calculate the ratio b_2/b_1 and the correction c for the different lengths of good and bad wires l_2 and l_1 in order to determine the amount of error arising from the assumption that $b_2/b_1 = l_1/l_2$. The values of capacitive reactance are determined from the fact that

$$b = Z \sin \phi,$$

where Z is the impedance and ϕ its angle as determined from formula (1).

The variation in b with variation in length of line is shown diagrammatically in Fig. 2. The total length of line is represented by Ol_2. The reciprocal of b is plotted on the vertical axis for different lengths of line. For an open at l_1 the location indicated by the ratio b_2/b_1 is at n whereas the true location is at m. The correction $mn = c$ must be subtracted from the apparent location to give the true location of the open. In the lower curve, this error is plotted against the total length of line l.

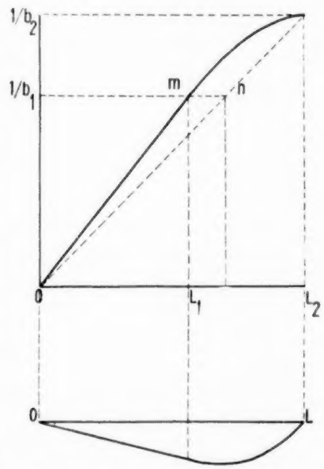


Fig. 2—Diagram showing the construction of a correction curve. Abscissas represent true linear distance; ordinates of upper curve represent measured capacitance; ordinates of lower curve represent errors of computed location.

It remains, then, to calculate b for a large number of lengths from zero to the maximum length of cable to be encountered. These values of b are used to calculate b_2/b_1 for different total lengths of line l and different fault locations l_1: that is, b is calculated for different lengths up to one hundred miles; then a set of ratios of b_2/b_1 and l_1/l_2 can be determined using the b of one hundred miles as b_2 and b for all the shorter lengths as b_1. Similarly, a set of ratios can be calculated for 95, 90, 85, 80, 75, etc., miles as total lengths. Since the interpolation of hyperbolic functions is at best a tedious calculation, even values of hyperbolics can be used in formula (1) and the corresponding odd lengths in miles calculated.

The ratio b_2/b_1 is then subtracted from the ratio l_1/l_2, in each instance this procedure resulting in a family of correction curves expressed in percentage such as that shown in Fig. 3. In this figure the correction is plotted against apparent rather than actual percentage distance in

order to facilitate the use of the curves in locating actual cases of trouble. Where the length of line being measured does not correspond to one of the curve indices, the curves can be interpolated and the desired curve drawn in.

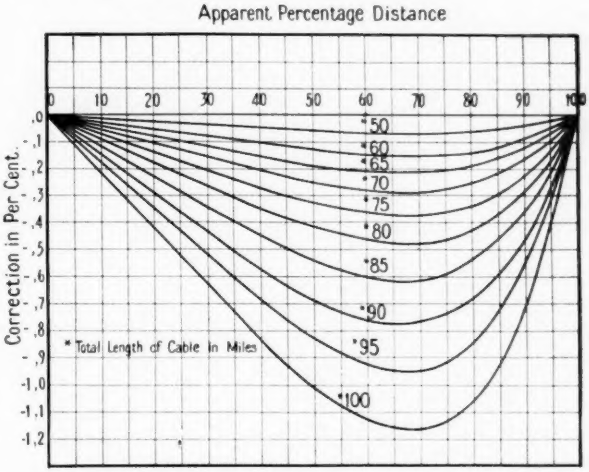


Fig. 3—Corrections to be used in finding true position of a fault from the apparent position. Apparent distance to the fault is indicated as a percentage of the total length of the cable.

The treatment becomes more involved for cables which are not homogeneous. The quads in the outer layers of any unspliced length of cable tend to differ in resistance from pairs in the inner layers, both because of greater length of turn in the outer layers and because of a possible difference in temperature between inner and outer layers. For homogeneous cables treated above these variations in resistance as well as attendant differences in capacitance to ground are equalized by mixing the quads among the various layers at each splice. In general, cables are spliced in this way, although there are several long distance cables now in service which are not homogeneous in this respect. Such cables are spliced in a special arrangement which is essentially a "transposition" system of splicing.

The "transposed" cable may readily be visualized by a consideration of the system of splicing which is employed. Instead of the 19-gauge quads being mixed promiscuously among all the layers, they are divided into an outer layer group and an inner layer group. The quads of each group are mixed among themselves (but not with

quads of the other group), at every splice but one. This particular splice, near the middle of the cable, is known as the "transposition point." At this splice the two groups are "transposed," that is, outer layer quads are spliced to inner layer quads and inner layer quads are spliced to outer layer quads. In this way the differences in resistance and capacitance to ground of outer and inner layer quads are averaged at the "transposition point" for each group. The average resistance or capacitance to ground of a conductor of the outer layer group will therefore differ appreciably from the average resistance or capacitance to ground of a conductor of the inner layer group. The constants given in Table 1 are for a cable of this type.

As in the case of the non-transposed cable, it is necessary to calculate values of b for different lengths of line. Up to the transposition point the procedure is the same as above, viz.,

$$Z_{l1} = Z_{O1} \coth P_1 l_1,$$

where the subscript denotes the first section adjacent to the measuring station. As soon as the point of open falls on the distant side of the transposition point, where the conductor changes layers, the calculation of Z_l is a composite one. That is

$$Z_{l12} = Z_{O1} \tanh (P_1 l_1 + \delta), \tag{2}$$

where Z_{l12} is the combined input impedance, Z_{O1} is the characteristic impedance of the adjacent section, P_1 and l_1 its propagation constant and length respectively, and

$$\delta = \tanh^{-1} \frac{Z_{l2}}{Z_{O1}},$$

where Z_{l2} is the input impedance of the distant section calculated from the formula

$$Z_{l2} = Z_{O2} \coth P_2 l_2.$$

However, the calculation of formula (2) involves practical difficulties, and it is best reduced as follows:

Denoting $P_1 l_1$ as θ_1 and $P_2 l_2$ as θ_2,

$$Z_{l12} = Z_{O1} \tanh \left(\theta_1 + \tanh^{-1} \frac{Z_{l2}}{Z_{O1}} \right),$$

and expanding,

$$Z_{l12} = Z_{O1} \left\{ \frac{\tanh \theta_1 + \dfrac{Z_{l2}}{Z_{O1}}}{1 + \tanh \theta_1 \left(\dfrac{Z_{l2}}{Z_{O1}} \right)} \right\}.$$

But

$$Z_{l2} = \frac{Z_{O2}}{\tanh \theta_2},$$

whence, substituting,

$$Z_{l12} = Z_{O1}\left\{\frac{\tanh \theta_1 + \dfrac{Z_{O2}}{Z_{O1}} \cdot \dfrac{1}{\tanh \theta_2}}{1 + (\tanh \theta_1)\left(\dfrac{Z_{O2}}{Z_{O1}}\right)\left(\dfrac{1}{\tanh \theta_2}\right)}\right\}$$

$$= Z_{O1}\left\{\frac{\tanh \theta_1 \tanh \theta_2 + \dfrac{Z_{O2}}{Z_{O1}}}{\tanh \theta_2 + \tanh \theta_1 \dfrac{Z_{O2}}{Z_{O1}}}\right\}$$

$$= Z/\phi.$$

Here, as in the case of the non-transposed cable, even values of hyperbolic functions are chosen and the corresponding odd values of length are calculated. Thus, the impedances of different arrange-

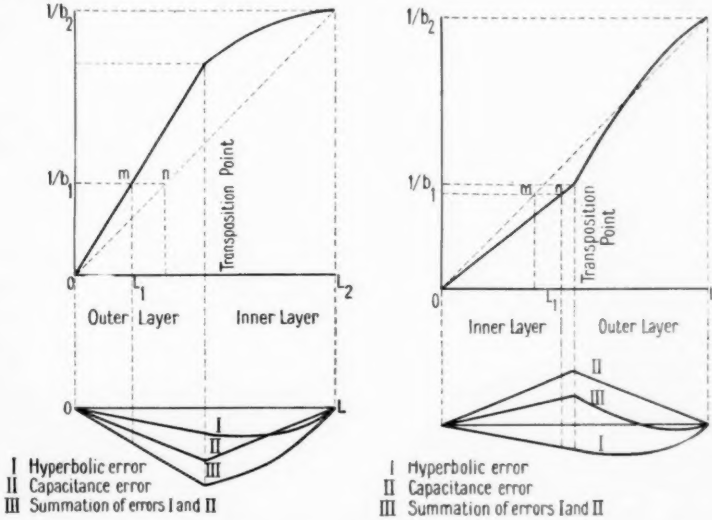


Fig. 4—Diagram showing the construction of a correction curve for a transposed cable. The fault is assumed to be at point L_1. Ordinates are the same as in Fig. 2. The lower curves I and II show the two errors which are algebraically added to give the total error.

Fig. 5—This diagram is similar to Fig. 4 except that the fault at L_1 is in that group of conductors which enters the measuring office in the inner layer of the transposed cable.

ments of conductor for a given transposed cable can be found and the calculated values of b_2 and b_1 may be used to plot a correction curve.

If, for a given transposed cable, a correction curve is calculated as outlined above, it will be found to have the general characteristics shown diagrammatically in Figs. 4 and 5. Fig. 4 is for the case where the faulty conductor enters the measuring station in the outer layer and Fig. 5 for the case where the faulty conductor enters the measuring station in the inner layer. In the lower half of each figure the total errors are separated into their component parts. In either Fig. 4 or Fig. 5 the total error curve can be assumed to be made up of two factors, a hyperbolic error similar to that shown in Fig. 2, and a straight line error due to the fact that the two halves of the cable do not have the same unit capacitance. The latter error would be present in such a cable even though its length were insufficient to cause a hyperbolic error. Such a division can be made because the constants of the inner and outer layers do not differ enough to affect the hyperbolic error appreciably. It is not possible to plot a general family of curves of the type shown in Fig. 3 due chiefly to the fact that the location of the transposition point and the difference in total length of different cables constitute a double variable. The need for a correction involving the double variable has been met by the development of open location equipment and methods which reduce this error to a negligible magnitude for the lengths and types of lines encountered in practice. The rigid treatment is, however, that outlined in formula (2).

The amount of the hyperbolic error can be calculated closely enough using the average constants of the inner and outer layers. The size of the straight line error due to the different capacitances of the inner and outer layers is found as follows:

Let C_1 and C_2 represent the adjacent and far end capacitances per unit length and D_1 and D_2 the respective lengths of these sections. The total conductor capacitance is then $D_1C_1 + D_2C_2$. If D is the location of the fault and D is less than D_1, that is, the fault is in the half of the cable adjacent to the measuring station, the bad wire capacitance is DC_1. The apparent location is

$$\frac{DC_1}{D_1C_1 + D_2C_2}$$

and the correction is

$$\frac{D}{D_1 + D_2} - \frac{DC_1}{D_1C_1 + D_2C_2}.$$

Similarly when the trouble occurs beyond the transposition point,

and D is greater than D_1, the capacitance of the bad wire is

$$\frac{D_1C_1 + (D - D_1)C_2}{D_1C_1 + D_2C_2}$$

or

$$\frac{D_1(C_1 - C_2) + DC_2}{D_1C_1 + D_2C_2}$$

and the correction is

$$\frac{D}{D_1 + D_2} - \frac{D_1(C_1 - C_2) + DC_2}{D_1C_1 + D_2C_2}.$$

The relative sizes of the capacitance and hyperbolic errors are shown in Fig. 6 where these two components and their sum are plotted as corrections against the apparent percentage distance to the fault.

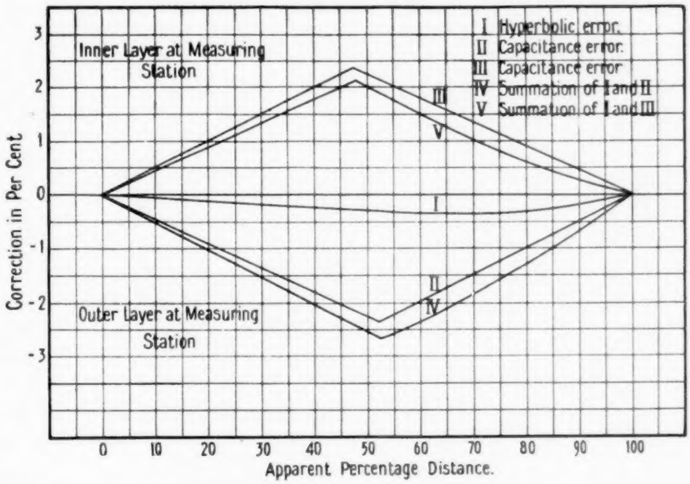


Fig. 6—Diagram for a transposed cable, showing the hyperbolic error as compared to the capacitance error,—measuring frequency four cycles.

In the development of a suitable open location method, it was necessary to select a definite frequency of testing potential for use with an impedance bridge. It was desirable that the selected frequency of testing potential should permit of a design of testing equipment which would have convenient operating characteristics. It was also desirable that the frequency of testing potential be selected to minimize errors which varied with frequency. The calculation of

hyperbolic errors for different frequencies of testing potential showed that this error decreased with frequency, the optimum value of frequency being zero. This relation is shown in Fig. 7, where the maximum errors at different frequencies for a 60-mile length of 19-gauge, non-loaded cable are plotted. However, with zero frequency, the sensitivity to unbalance for an impedance bridge network is also zero, increasing as the frequency increases.

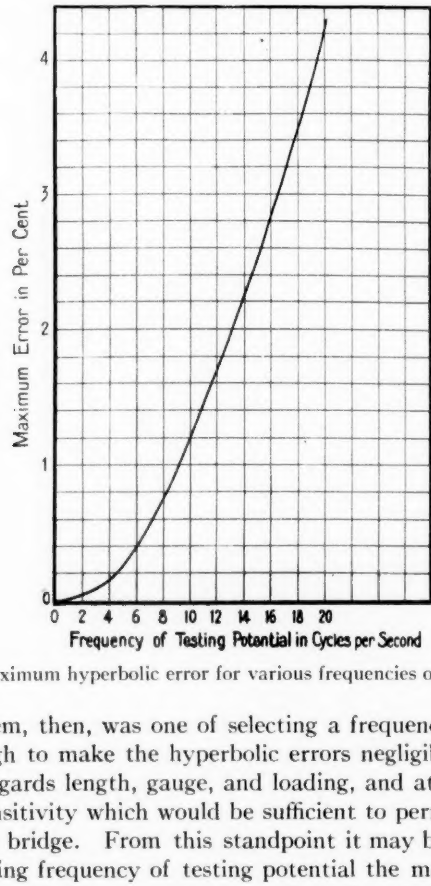


Fig. 7—The maximum hyperbolic error for various frequencies of testing potential.

The problem, then, was one of selecting a frequency which would be low enough to make the hyperbolic errors negligible for all cases of lines as regards length, gauge, and loading, and at the same time provide a sensitivity which would be sufficient to permit an accurate balance of a bridge. From this standpoint it may be observed that for a decreasing frequency of testing potential the maximum rate of decrease of hyperbolic error appears at about four cycles as shown by Fig. 7. A computation of the hyperbolic error for measurements made at four cycles on sixty miles of cable gives results as follows:

Type of Cable Circuit	*Average Percentage Error*
Extra Light or Non-Loaded 19-Gauge Cable	0.083%
Medium Heavy or Heavy Loaded 19-Gauge Cable	0.033
Extra Light or Non-Loaded 16-Gauge Cable	0.022
Medium Heavy or Heavy Loaded 16-Gauge Cable	0.090

Since these errors, at a frequency of four cycles, are for the maximum length of line which may be encountered in practice, it was considered that they might be neglected in comparison with the importance of securing a frequency of testing potential which would be high enough to give an impedance bridge a suitable sensitivity to unbalance. Hence, a computation of the sensitivity appeared to be the next step in the selection of a suitable frequency of testing potential.

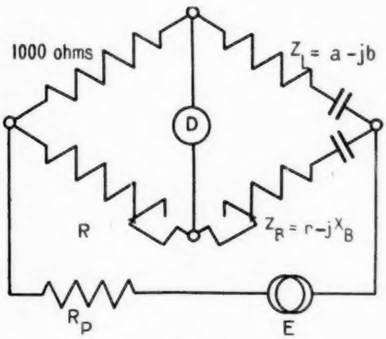


Fig. 8.

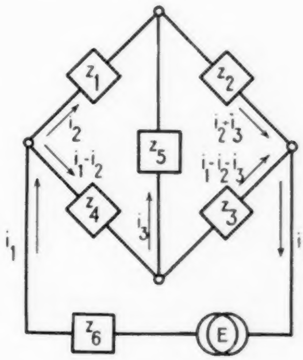


Fig. 9.

Fig. 8—Bridge used in the location of opens. E represents the low-frequency source, and R_p a protective resistance. The rheostat R and the 1000-ohm resistance may be regarded as the ratio arms of the bridge. The impedance of the line is represented by Z_L, a resistance and capacitance in series. The rheostat r is used in balancing the resistance component of the line impedance, so that the impedance angle of Z_B will equal the impedance angle of Z_L.

A very sensitive electrodynamometer, or a galvanometer equipped with an electromagnetic field, was used as a detector in the impedance bridge network shown in Fig. 8. The sensitivities for several frequencies of testing potential were computed for this modified form of the De Sauty bridge. The condition for balance in this impedance bridge network is

$$RZ_l = 1000Z_B$$

or

$$R(a - jb) = 1000(r - jX_B).$$

Collecting reals and imaginaries,

$$Ra = 1000r$$

and

$$Rb = 1000X_B.$$

If measurements R_1, r_1 and R_2, r_2 are made on the bad and good wires respectively,

$$R_1a_1 = 1000r_1$$

and

$$R_2a_2 = 1000r_2;$$

also

$$R_1b_1 = 1000X_B$$

and

$$R_2b_2 = 1000X_B,$$

$$R_1b_1 = R_2b_2,$$

or

$$\frac{R_1}{R_2} = \frac{b_2}{b_1}.$$

Yet the design of a suitable bridge is not concerned alone with balanced condition, but rather with the sensitivity and ease of balance with slight unbalances present. An indication of the probable sensitivity is afforded by a solution of this bridge network to determine the phase and magnitude of the galvanometer unbalance current with respect to the impressed voltage. These have been obtained from the equation

$$i_3 = \frac{E(z_2z_4 - z_1z_3)}{\begin{vmatrix} -z_5 & (z_1 + z_4) & -z_4 \\ (z_2 + z_3 + z_,) & (z_2 + z_3) & -z_3 \\ -z_3 & -(z_3 + z_4) & (z_3 + z_4 + z_6) \end{vmatrix}} \tag{3}$$

in which the denominator is a determinate, and the symbols are those used in Fig. 9.

Since the condition for balance is

$$z_1z_3 = z_2z_4$$

we have, using the notation of Fig. 8,

$$1000(r - jX_B) = R(a - jb)$$

and, equating reals and imaginaries,

$$1000r = aR, \tag{4}$$

$$1000X_B = bR. \tag{5}$$

This impedance bridge can be balanced in two ways: by varying r and X_B, keeping R constant, or by varying R and r, keeping X_B constant (Fig. 8). The effect is essentially the same in either case. When R is varied instead of X_B, the only difference is that the balance of the bridge is disturbed for both the real and imaginary components. This fact necessitates a correction of r each time R is changed in securing a balance of b against X_B. If X_B is varied, the balances of r against a, and X_B against b, are independent functions. In practice, it is easier to vary R and keep X_B constant, but for the purpose of theoretical discussion it lends clarity to consider X_B to be variable from the condition of balance.

From the equation (1) above the impedances of different lengths of line may be computed. For the purpose of designing a suitable impedance bridge arrangement, it is sufficient to consider the 19-gauge, non-loaded cable only, as the effect of loading on the general line characteristic is small at the frequencies employed. A number of impedance values representing different lengths of 19-gauge, non-loaded

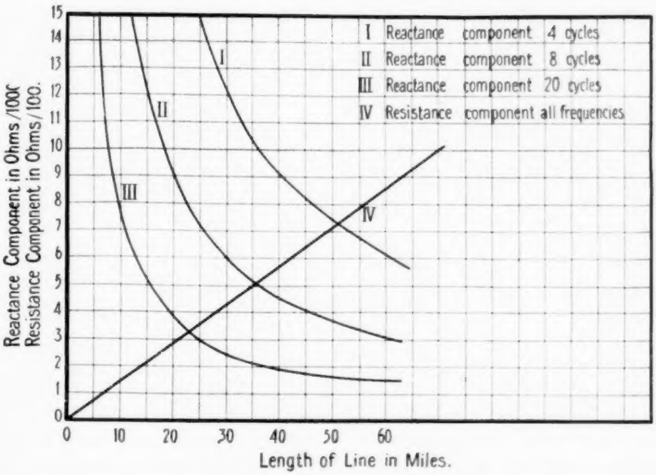


Fig. 10—Characteristics of the resistance component and capacitance component of the impedance of lengths 19-gauge, non-loaded cable.

cable up to sixty miles, at three frequencies, viz., twenty, eight and four cycles, were selected and the condition of balance of the impedance bridge calculated for each case from equations (4) and (5). Curves of these impedances are shown in Fig. 10, where the reactances and resistances at the three chosen frequencies are plotted.

Since the sensitivity of the bridge network should be a maximum when the unbalance is small, i.e., when a balance is about to be secured, this condition is the one with which the sensitivity calculation is concerned. The capacitance X_B of the bridge network was assumed to vary 10% from the condition of balance and the galvanometer

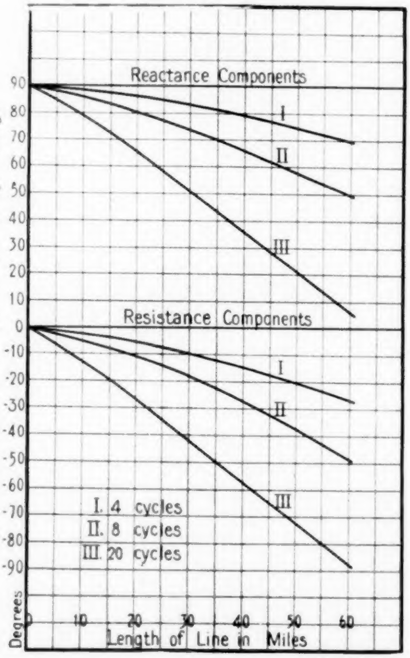


Fig. 11—The variation of the angle between the bridge voltage and the unbalance current for the reactance component and resistance component of non-loaded 19-gauge cable in lengths varying from zero to sixty miles. The curves represent characteristics for frequencies of four, eight and twenty cycles.

unbalance current, i_3, was then calculated from equation (3) for each length of line at each frequency. Both the magnitude and phase angle were found. Similarly the resistance r was assumed to vary 10% from the condition of balance, X_B remaining balanced, and another set of galvanometer unbalance currents was determined. Such calculations are particularly tedious, involving successive additions and subtractions, multiplications and divisions of complex quantities. The results of these calculations are shown in Figs. 11, 12 and 13. Fig. 11 represents the variation in phase angle of i_3 with

respect to the bridge potential, Fig. 12 the magnitude of i_3, and Fig. 13 the relative sensitivities obtained with different frequencies and different lengths of line. These sensitivities are proportional to i_3 (Fig. 12) and to the cosine of the angle between i_3 and the field current

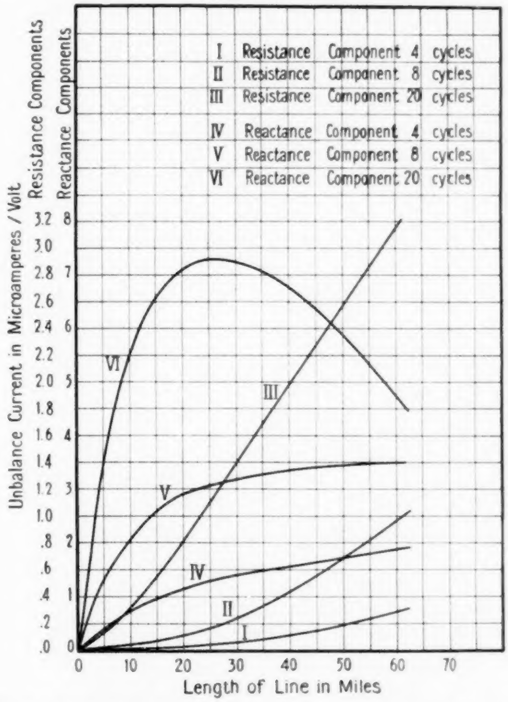


Fig. 12—Magnitude of the unbalance current for measurements, on the impedance of non-loaded 19-gauge cable, at frequencies of four, eight, and twenty cycles.

of the alternating-current galvanometer. In the calculation of the curves of Fig. 13 the field current of the galvanometer was assumed to be in phase with the bridge potential for the case of an unbalance in r, and leading the bridge potential by ninety degrees for the case of an unbalance in X_B.

Referring to Fig. 11, it is seen that for short lengths of line the unbalance current caused by unbalancing r is almost in phase with the voltage, while the unbalance current due to unbalancing X_B leads the voltage by approximately ninety degrees. As the length of line increases, the phase angles tend to lag from these positions, due to

the effect of the convergent variation of the resistive and reactive components of the line impedance, that is, the resistance increases and the reactance decreases with increase in length of line. This lag is greater for the higher frequencies. The total variation for sixty miles of cable measured at twenty cycles is practically ninety degrees which means that for a given field current the sensitivity using twenty cycles must approach zero with some length of line between zero and sixty miles. This condition is illustrated in Fig. 13, where the sen-

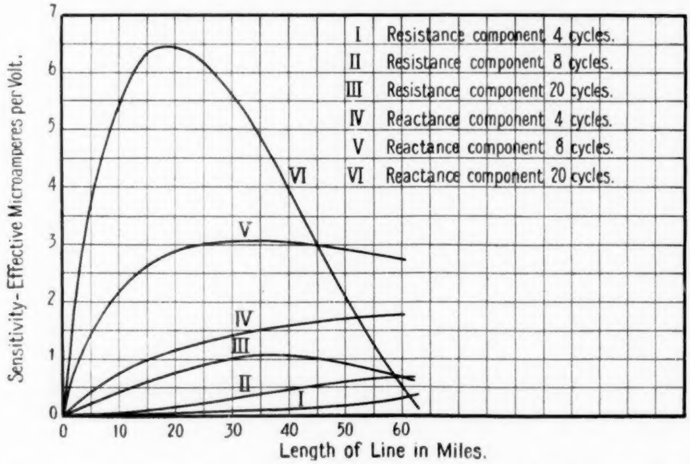


Fig. 13—Sensitivity to unbalance for the bridge network for impedance measurements on lengths of non-loaded 19-gauge cable. These curves represent sensitivities for frequencies of testing potential of four, eight and twenty cycles.

sitivity for twenty cycles is a maximum at twenty to thirty miles, but falls rapidly toward zero at the longer lengths of line. With the lower frequencies and the setting of field current used, the general effect is an increase in sensitivity as the length of line increases, and a decrease in sensitivity with decrease in frequency. This decrease is due to the decrease in reactance with increase in length of line (Fig. 10).

It would appear that provision should be made for shifting the phase of the field current through ninety degrees, its two positions being respectively in phase with the bridge potential and ninety degrees leading. Thus the two components, r and X_B, could be balanced independently except for the shift in phase with different line lengths. This phase shift is small with a frequency of testing potential of four cycles.

A frequency of four cycles was selected as the optimum as regards the size of hyperbolic error discussed above, the sensitivity available, and the amount of phase shift with increase in length of line. The curve showing the change of hyperbolic error with variation in frequency (Fig. 7) shows four cycles to be at or near the critical point of the curve. The sensitivity at four cycles has the advantage of being sufficient but not excessive. To a large extent, the condition of phase shift (with increase in length of line) governs the ease of securing a balance over the range of line lengths. The ideal arrangement would be one in which the field current could always be placed in phase with the component of the bridge unbalance current it was desired to eliminate. Such a quality is not characteristic of the type of bridge used; however, a desirable approximation of such an arrangement is obtained when a four-cycle frequency of testing potential is used.

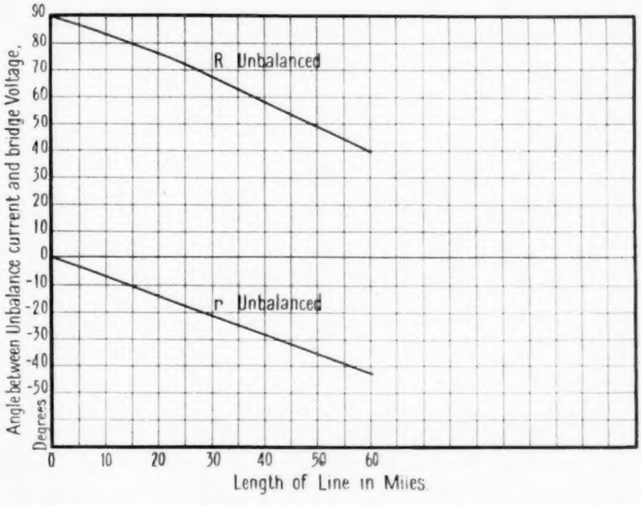


Fig. 14—Calculated variation in the angle between unbalance current and bridge voltage for unbalance in the capacitance component R and for unbalance in the resistance component r of the impedance of non-loaded 19-gauge cable. Testing potential of 4-cycles.

In order to check the assumption (stated above) that the bridge could also be balanced by keeping X_B constant and varying R and r, without materially changing the conditions of balance, another set of galvanometer unbalance currents was calculated with X_B constant and R and r varied respectively 10% from the condition of balance.

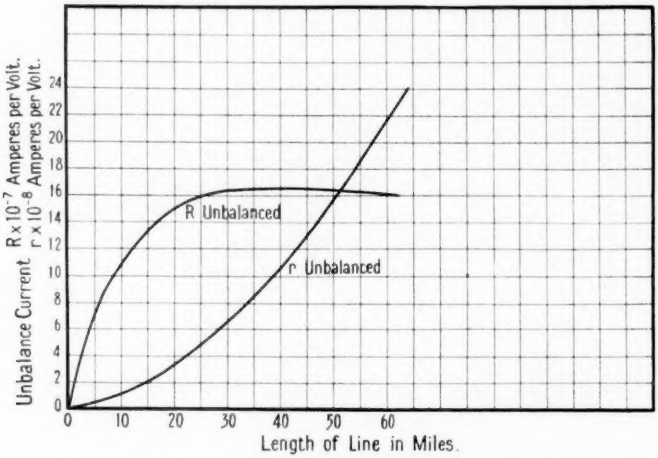


Fig. 15—Calculated magnitude of the unbalance current for unbalance in the capacitance component R and unbalance in the resistance component r of the impedance of non-loaded 19-gauge cable. Testing potential 4-cycles.

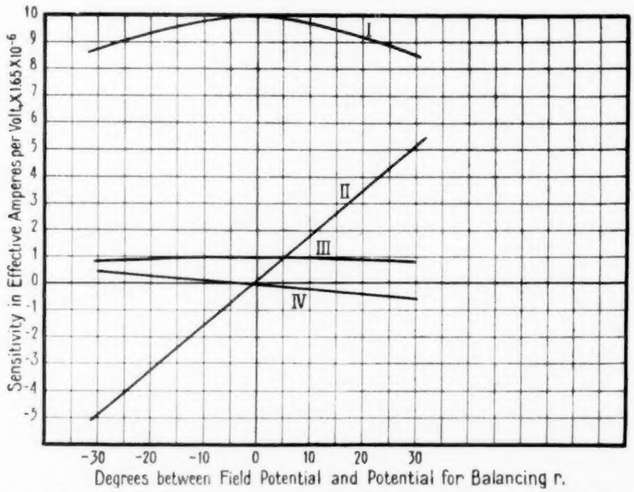


Fig. 16—Calculated sensitivity of the bridge at four cycles on a 50-mile length of non-loaded, 19-gauge cable. Sensitivities for unbalance in the component:

I—R with the phase of testing potential applied for balancing R.

II—R with the phase of testing potential applied for balancing r.

III—r with the phase of testing potential applied for balancing r.

IV—r with the phase of testing potential applied for balancing R.

The chosen frequency of four cycles was used in this calculation. These curves of phase angle and current magnitude are shown in Figs. 14 and 15, and correspond to those shown in Figs. 11 and 12, calculated by the other method.

Since R increases with increase in length of line, the sensitivity per ohm change in R will be better throughout the range of lengths if the sensitivity for a given change in R is a maximum when R is a maximum. Taking fifty miles as the average total length of line, and

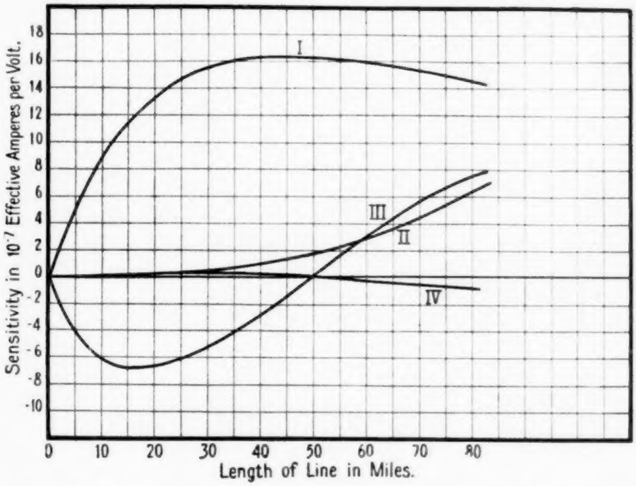


Fig. 17—Calculated sensitivity for lengths of non-loaded 19-gauge conductor using the impedance bridge arrangement which gave the maximum sensitivity for curve I in Fig. 16. Sensitivity for unbalance in the component:

I—R with the phase of testing potential applied for balancing R.
II—r with the phase of testing potential applied for balancing r.
III—R with the phase of testing potential applied for balancing r.
IV—r with the phase of testing potential applied for balancing R.

assuming two bridge potentials ninety degrees apart, a set of sensitivity curves was calculated for this length of line, the phase of the field potential being varied on either side of the bridge testing potential. The lag of the field current from the field voltage calculated from the field inductance and resistance was found to be about forty degrees. The resulting sensitivity curves are shown in Fig. 16, and these indicate a maximum sensitivity for R and r with the field potential in phase with the potential used to balance r.

With the assumed conditions as determined by this calculation,

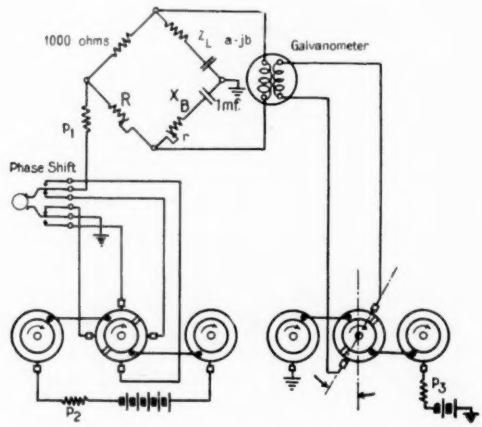


Fig. 18—Impedance bridge circuit showing the arrangement used in applying testing potentials in quadrature.

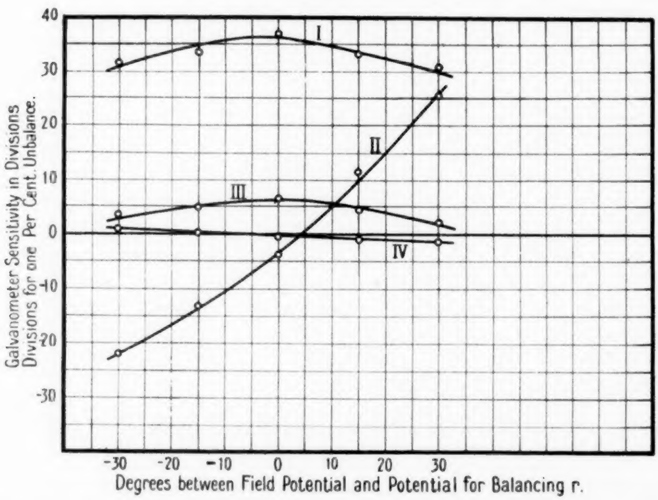


Fig. 19—Observed galvanometer sensitivities which are comparable to the calculated sensitivities of Fig. 16. Sensitivity for unbalance in the component:

I—*R* with the phase of testing potential applied for balancing *R*.
II—*R* with the phase of testing potential applied for balancing *r*.
III—*r* with the phase of testing potential applied for balancing *r*.
IV—*r* with the phase of testing potential applied for balancing *R*.

viz., two bridge potentials ninety degrees apart and a field current lagging one bridge potential forty degrees, a complete set of sensitivity curves was calculated for different lengths of line from zero to eighty miles. These are shown in Fig. 17. It should be noted that the sensitivity for detecting an unbalance in R is a maximum at the desired length of fifty miles. The sensitivity for an unbalance in r

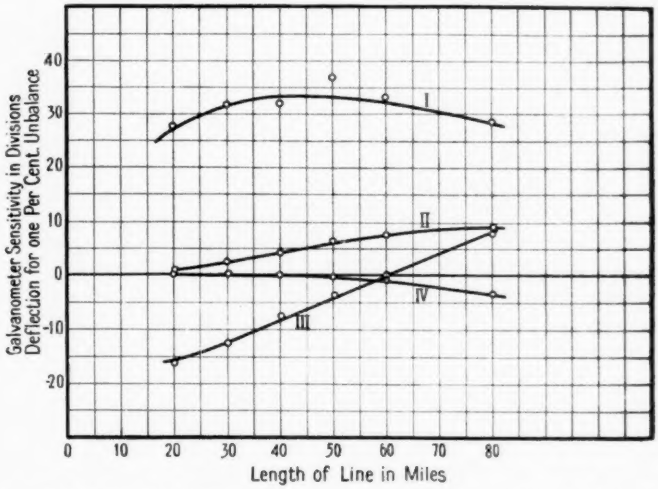


Fig. 20—Observed galvanometer sensitivities which are comparable to the calculated sensitivities of Fig. 17. Sensitivity for unbalance in the component:

I—R with the phase of testing potential applied for balancing R.
II—r with the phase of testing potential applied for balancing r.
III—R with the phase of testing potential applied for balancing r.
IV—r with the phase of testing potential applied for balancing R.

is low at the shorter lengths of line, but increases as the length of line increases. It may be noted that in both Figs. 16 and 17 the sensitivity curve for changes in R passes through zero when the testing potential is applied for balancing r. Likewise the sensitivity curve for changes in r passes through zero when the testing potential is applied for balancing R. This point is where the field current and galvanometer unbalance current are ninety degrees out of phase. Naturally this point coincides with the point of maximum sensitivity for the normal potential arrangement. As stated above, it would be ideal if these "reverse sensitivities" could be zero and the "true sensitivities" could be maximum throughout the entire range of lengths. Reference to Fig. 17 will show that the reverse sensitivity

for r is practically zero throughout the range of lengths. This fact in itself is significant. Assuming r to be set on zero, R could be varied to secure an approximate balance, using the proper testing potential. This balance would be fairly accurate since the reverse sensitivity for r is quite low throughout. Shifting bridge potentials ninety degrees, r could be adjusted almost to the proper point since R is practically

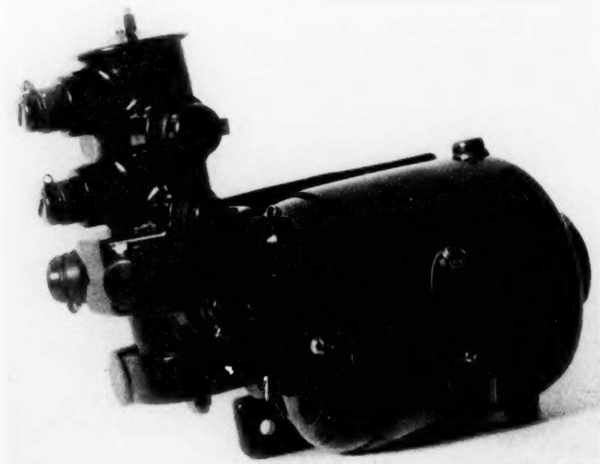


Fig. 21—Commercial design of motor-driven commutator which is used to supply 4-cycle alternating potentials for the impedance bridge.

correct. The balance of R and r can then be refined as often as is necessary to secure a perfect balance. Since r is not used in any calculation, and since its effect on R is small once an approximate balance of R and r is obtained, the need for an accurate balance of r is small.

A series of observations made with experimental apparatus arranged as shown in Fig. 18 gave the sensitivity curves shown in Figs. 19 and 20. The observed sensitivity characteristics shown by these curves agree very favorably with the theoretical values shown by the curves of Figs. 16 and 17.

By way of summarization, it may be observed that in the process of developing a suitable method for the location of faults in telephone cables a definite sequence of steps has been taken to provide an effective treatment of the problem:

1. In establishing requirements for a suitable method, a study was

made of the historical development of the art. The effectiveness of the art was compared with the needs of present practices.

2. Preliminary to the development of a suitable method, an analysis was made of the errors which may enter into the determination of an open location. These errors were classified for treatment or correction during the development.

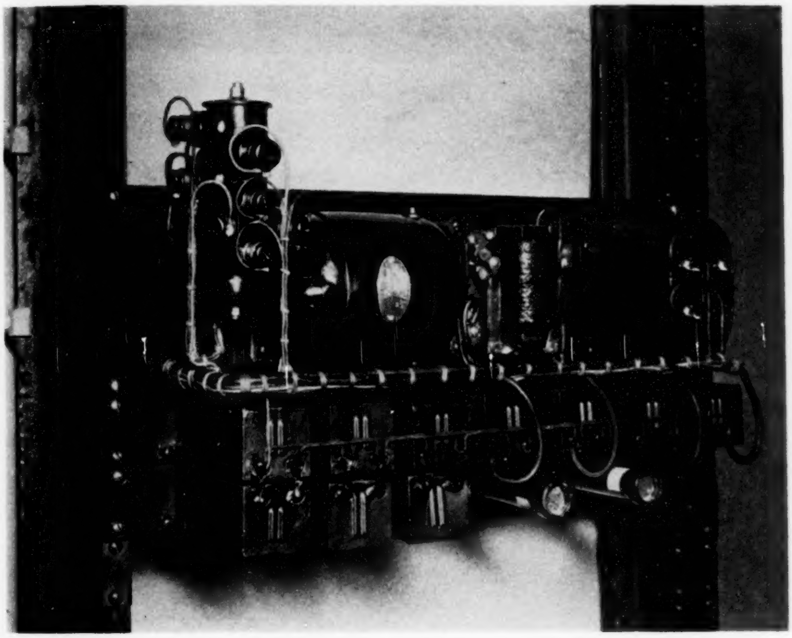


Fig. 22—4-cycle motor-driven commutator with associated equipment.

3. A mathematical analysis was made of the electrical characteristics of the circuits on which open locations may be required.

4. A suitable method was devised and an associated impedance bridge circuit was developed which, in the light of the recent research, most consistently and economically met the requirements for the determination of open locations.

After having completed an analysis of the problem and demonstrated the practicability of the proposed methods, it remained to develop applications of these methods for practical use. Equipment was designed to develop a low frequency source of alternating potential

by reversing a testing battery. The device developed for this purpose is a 4-cycle, motor-driven commutator shown in Fig. 21. By studying the curves used in the selection of a suitable frequency of testing potential, it may be observed that the selected value of four cycles is not critical, in fact a variation of $\pm$ 25 per cent may be allowable in

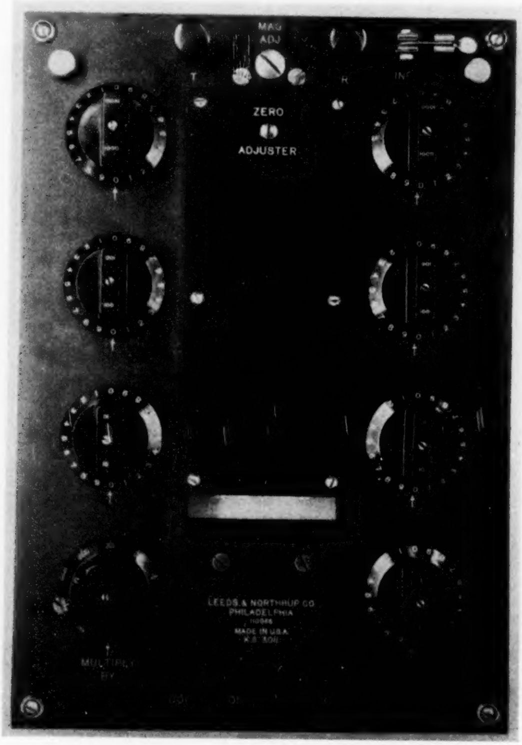


Fig. 23—Panel assembly of the impedance bridge.

different machines. However, in practical application, the accuracy of results obtained depends upon a comparison or ratio of two impedance measurements which cannot be made simultaneously. Since impedance varies with frequency, it is important that the frequency of testing potential should remain constant while the two measurements are being made. This requirement is met for the short time required for two measurements. The assembly of the 4-cycle commutator with associated apparatus is shown in Fig. 22.

The galvanometer and the equipment required in the modified form of the De Sauty bridge have been incorporated in a compact unit which is shown in Fig. 23. This bridge, while being particularly adapted to the 4-cycle impedance measurements required for open location tests, is also applicable for direct-current bridge measurements. Assembly details for the bridge arrangement are shown in Fig. 24.

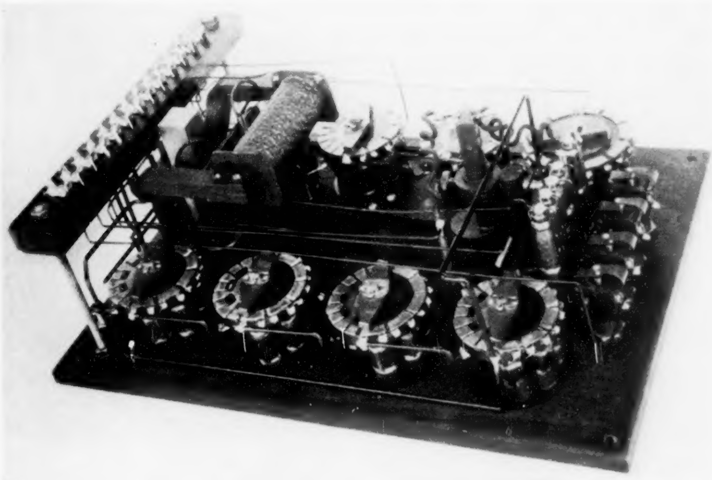


Fig. 24—Internal assembly of the impedance bridge.

The alternating-current galvanometer is sufficiently sensitive to be directly actuated by any significant unbalance of the impedance bridge, so that it is unnecessary to use an amplifier, rectifier, or other converting apparatus which may be difficult of adjustment or maintenance.

A few additional features are outlined as some of the significant results of this development of an improved open location method.

It has been shown that the error, caused by the deviation from the straight line relation between sending end admittance and physical length of line, has been reduced to a value which may be neglected as being less than the required accuracy of the open location method.

As a result of an analysis of errors which are introduced by line irregularities, methods were devised for applying corrections for all such errors which are of a magnitude sufficient to interfere with the desired accuracy.

This impedance bridge arrangement employs several features which are a distinct improvement on the methods previously utilized for this purpose. In order that the impedance angles of the impedance networks may be balanced, a variable resistance has been connected in that branch of the bridge which contains the comparison capacitance. This balance of the impedance angles of the impedance network was found to be important in obtaining a steady performance of the alternating-current galvanometer. A system has been devised for separately balancing the resistance components as well as the capacitance components of the reactive networks of the impedance bridge. It was found that this arrangement gives a maximum sensitivity to each component and permits of a very rapid balance of the bridge.

Contemporary Advances in Physics—XII. Radioactivity

By KARL K. DARROW

IN the year 1896, which fell near the beginning of the great transformation of modern physics, Henri Becquerel heard that Roentgen had discovered strange rays proceeding from an electric discharge-tube while the discharge was passing and the glass walls of the tube were phosphorescing. Suspecting that the new rays were connected with the phosphorescence, Becquerel tested samples of some of the substances which naturally phosphoresce. It happened that one which he tested was a compound of uranium. He wrapped the sample in paper to shut in the light of its phosphorescence, and set it beside a photographic plate; for the rays of Roentgen had disclosed themselves by acting on such plates. Becquerel had made a happy guess; for the compound affected the plate. Yet his original idea was altogether wrong; for the effect had nothing to do with the phosphorescence of the compound, it was due to the uranium itself and faithfully reappeared when other and non-phosphorescent compounds were used instead, and even when a piece of the pure metal was set beside the plate. It was an instance of a fallacious idea having guided a keen observer to a great discovery—not the first in the history of physics, and assuredly not the last.

Thereupon Pierre and Marie Curie, having verified that the effect of any quantity of any compound of pure uranium is strictly proportional to the amount of uranium in it, noticed that the effect of certain natural rocks and minerals containing uranium was much greater than that which their content of the metal should produce. Suspecting that there was some constituent of the rocks having the same property as the uranium but in a degree much greater, they set about the task of getting the uranium and the inert matter out of the way and isolating the more potent substance. It was a long task; to speak of "winnowing" the pile of rock would be to use a comically feeble metaphor, and as for the proverbial needle in a haystack, it could have been extracted with incomparably less trouble than the few hundredths of a gramme of the active substance which were latent in the ton of raw material. Eventually the Curies did liberate it, or rather them, for there were several active substances; and one of them was named radium, and their strange property was called radioactivity. This was the first of the words containing the magic syllables *radio*, syllables which are one of the special symbols of our epoch; were the

literature of these times to disappear all but a few scraps, posterity could date them by the appearance of that word, as Latin manuscripts are dated through containing some word or some trick of style that came into use at a definite moment of history. A word must bear almost magical connotations, to enter so thoroughly into popular usage; and the phenomena of radioactivity endowed it with these in abundance, with suggestions of rays piercing all matter, and inexhaustible stores of energy, and transmutation of elements, and influences having power even over life and death. Wonderful to relate, the suggestions for once were justified by the truth.

From 1898 onward there was a tremendous rush of investigators into the new field, and in a few years there were explorers of almost every conceivable aspect of radioactivity—chemists ascertaining the chemical properties of the radioactive elements, physicists observing their physical properties, and a great host of students investigating the numerous and striking effects of the rays. The subject presently became so wide that books on radioactivity written before and during the War resemble treatises on the contemporary physics of their dates of publication; for the new rays seemed to be able to invade all the provinces of physics as easily as they could penetrate matter in all its forms.

Eventually, however, it became clear that many of the topics classed at first with radioactivity should be removed into other fields of science. The radioactive elements all have their places in the Periodic Table, and their chemical properties are what should be expected from elements thus placed; peculiar as radium is in its one famous feature, there is nothing abnormal about its chemical reactions, and they may justly be relegated to the handbooks of chemistry and to the manuals written for those who wish to prepare or purify the element. The same thing is true of the physical properties of radium; nothing in its optical spectrum suggests that it is other than an ordinary member of the second column of the Periodic Table, nothing in its X-ray spectrum intimates that it is more than just the 88th member of the Procession of the Elements. None of these needs to be taken into account in the study of radioactivity.

The various effects of the rays which the radioelements emit are likewise quite irrelevant. At the beginning it was natural and proper for every writer to describe all that was known of the actions of the alpha-rays, the beta-rays and the gamma-rays, after having said that these are the three kinds of rays which radioactive substances emit. Indeed it was essential, for at first there was no way of defining the rays, much less of ascertaining their real nature, except by considering *en bloc*

everything that was known about their actions. This condition prevails no longer. It is established that alpha-rays are atoms of helium each bearing a charge $+2e$; that beta-rays are electrons, that gamma-rays are composed of electromagnetic radiation. Information about the first two belongs to the vast body of doctrine concerning the properties of fast-flying electrified particles; information about the last belongs to the science of the properties of radiation. I do not mean to imply that the information is redundant. One can produce in the laboratory fast-flying electrons, but none so fast as the fastest beta-rays; swift positively-charged atoms, but none nearly so swift as the alpha-particles; electromagnetic waves of many wavelengths, but none nearly so short as the shortest to be found among gamma-rays. The knowledge acquired from studying the properties of the rays is exceedingly important, and if the radioelements had not been discovered, most of it would not been acquired so early, and some of it would still be unattainable; but it is not knowledge of radioactivity.

What then is knowledge of radioactivity? So far as now appears, we know all that can be known about the radioactivity of a radioelement if we know what are the speeds of the alpha-particles emitted from it, if any; what are the speeds of the electrons emitted from it, if any; what are the wavelengths of the electromagnetic waves which it emits, if any; how many of each kind of particle (for we may speak of the waves as particles also, meaning by "particle" a quantum) are emitted from a given number of atoms in a given time; and what element or elements result from these processes. Apparently, if we could know all of these things for a particular radioelement, we should know everything which determines its peculiar actions upon the outside world. This unfortunately is not the same thing as being able to solve the problems of predicting all of these actions or understanding them; but these problems are now transferred out of the field of radioactivity into the field of the science of fast-flying charged particles and short-wave radiation. Let us leave them there, and restrict the field of radioactivity to the speeds of the particles and the frequencies of the waves which issue from each radioelement, and the rates at which they come forth, and the condition of the atoms they leave behind.[1]

[1] The specific statements made in this article are derived chiefly from three recent synopses of the data of radioactivity: the National Research Council bulletin *Radioactivity*, by A. F. Kovarik and L. W. McKeehan; the *Manual of Radioactivity*, by G. v. Hevesy and F. Paneth; and the relevant articles by St. Meyer, L. Meitner, W. Bothe and O. Hahn in volume 22 of the new Geiger-Scheel *Handbuch der Physik*. As these are all well supplied with bibliographies (and so likewise, I presume, is the new edition of Meyer and von Schweidler's *Radioaktivität*) I have omitted references to individual papers except a few published since 1923. At several places I venture to refer under the name " Introduction " to my *Introduction to Contemporary Physics* for topics not falling within the field of radioactivity as here defined.

During the composition of this article I have had the advantage of frequent consultation with my colleague Dr. L. W. McKeehan.

Already in expressing these restrictions, certain principles of radioactivity have been implicitly assumed; it is necessary to state them explicitly.

In the first place, I have spoken of the radioactivity of the elements alone; this is permissible, because radioactivity is definitely a property of individual elements. This does not mean merely that radioactivity is a property of a limited number of elements in certain states and a limited number of compounds of these and other elements, as seems to be true of ferromagnetism. It means that wherever there is a particular radioactive element, free or compounded, gaseous or liquid or solid, the characteristic rays of that element are emitted in a degree proportional to the amount of the element and not affected in the least by its condition or its state of combination. A given amount of radium emits the same kinds of rays at the same rate whether it is a piece of pure metal, or is combined with chlorine in radium chloride, or with sulphur and oxygen in radium sulphate. A given amount of radon emits rays of the same sort at the same rate whether it is gaseous as at normal temperatures, or frozen by submerging its enclosing tube in liquid air. Samples of some of the radioelements have passed through combination after combination in the chemical laboratory, being released from one compound only to enter into another; their activity was meanwhile being measured by the most delicate available tests, but it was never found to be affected in any perceptible degree. There is no other property of an element, excepting mass, of which this can be said without reservation.[2]

The indifference of radioactivity to the state of combination of the elements which display it extends also to all their other circumstances. In modern laboratories it is feasible to subject pieces of matter to very powerful, severe and violent agencies; heat enough to melt any element, cold enough to freeze any substance, electric fieldstrength high enough to tear electrons out, high magnetic fields, intense illumination, bombardment by multitudes of fast moving charged particles—and all of these have been tried to some extent, some to the utmost humanly possible extent, upon radioactive elements; but in every instance the radioactivity has remained constant without detectable variation, inaccessible and immune to all the powers within human control or under human observation.[3]

[2] It can be said of the higher-frequency emission-lines and absorption-edges of the X-ray spectra of the elements, but not unreservedly; for since the lower-frequency lines and edges of an element do vary slightly but perceptibly when its state of combination is altered, there is a strong presumption that the higher-frequency spectra will likewise be found to vary as soon as the accuracy of the measurements is increased say five- or tenfold.

[3] Influences of sunlight upon radioactivity are reported now and then in the *Comptes Rendus;* but it seems exceedingly unlikely that something immune to every other known agency should be susceptible to this particular one.

Sooner or later, in expounding almost any topic in physics, one arrives at a place where the introduction of an atom-model greatly simplifies what remains to be said. In the present article, this is the place.

Physicists commonly employ an atom-model in which a certain number of electrons are arranged around a nucleus bearing a charge equal in magnitude and opposite in sign to the sum of their charges. For any particular element the number of electrons assigned to its atom-model is equal to its atomic number, which can be obtained from any modern chart of the Periodic Table. In such a chart the elements are arranged in the order of their atomic numbers from 1 to 92, composing what I shall call the *procession of the elements*—a procession from which only two are now missing. In dealing with an element of high atomic number—all of the radioactive elements are of this character, ranging in atomic number from 81 upwards [4]—the electrons are assigned to various locations, some being close to the nucleus and others intermediate and others at the periphery of the atom-model. In fitting the various regions and divisions of the atom-model to the various properties of the element which it represents, the outermost electrons are assigned to the task of accounting for those properties which vary exceedingly with the state of chemical combination and with the other circumstances of the element; for being at the surface of the atom they should be most exposed to outer influences. The inner electrons, being partly shielded, are used to account for such properties as the X-ray frequencies, which depend so little upon the circumstances of the element that their variations are scarcely perceptible or not at all. The nucleus is the best shielded of all, and it receives for its quota the two properties which within the accuracy of experiment are immune from change—radioactivity and mass.

There are additional reasons for assigning mass and radioactivity to the nucleus. As for the mass: since the sum of the masses of the electrons constituting an atom-model never attains 1/1800 of the known mass of the atom, the balance which the nucleus must take is practically the whole of it. Again, there are experiments which show that a single chemical element may have several kinds of atoms differing in mass and yet quite alike in chemical properties, in their line-spectra, in their X-ray spectra; since these similarities require that the same nucleus-charge and the same number and arrangement of electrons be imposed upon all these atoms, the outstanding difference in their masses must be ascribed to their nuclei.[5] Again, there are slight

[4] Except potassium and rubidium (compare footnote 13).

[5] *Introduction*, pp. 29–39, 65–66.

differences between the band-spectra of compounds involving such atoms, which are well explained by attributing to the several atoms identical nucleus-charges and electron-systems, but nucleus-masses standing to one another in the same ratios as the observed masses of the atoms do.[6] But I must not give all the evidence for the nuclear atom-model, or this article will be swamped.

Among the reasons for ascribing radioactivity to the nucleus, the primary one has already been introduced—radioactivity, like mass, is unalterable; and another has already been stated, though without mentioning its relevance to this question. Certain radioactive elements emit charged atoms of helium; and since outside of the nucleus nothing except electrons is provided in the atom-model, these charged atoms must be supposed to proceed out of the nuclei. This argument could not be used upon the radioelements which emit electrons; but even for these there are reasons for suspecting that some of the electrons which issue from them do not come out of the family of electrons surrounding the nucleus, but from some other place. For instance, it is possible and usual to pry electrons out of various locations in the circumnuclear family; but when this is done, the resulting "ionized" atoms promptly take in one electron or as many more as they have lost, and revert to their original state and nature. This does not happen with the radioactive atoms which emit beta-rays; the departure of the electron effects an irreversible change, the atom is altered for good and all. It does not however acquire a permanent positive charge; it takes on an electron and makes good its loss of charge. This is best explained by supposing that the original atom lost an electron originally located in the nucleus, and added one to the circumnuclear family, keeping its net charge equal to zero but undergoing a rearrangement of its charges.

By accepting the idea that certain of the charged particles emerging from a radioactive element issue from the nuclei of its atoms, it is possible to express and explain very simply a celebrated law of radioactivity which was discovered by Fajans and Soddy in the early days of the nuclear atom-model and helped greatly to establish it.

When an atom of a radioelement of atomic number Z emits an alpha-particle with its charge $+ 2e$, its nuclear charge diminishes by that amount. It becomes an atom with nuclear charge $(Z - 2)e$ and Z electrons. The diminished nuclear charge cannot hold the entire electron-family; two of its members depart, and the atom becomes an atom of nuclear charge $(Z - 2)e$ and $(Z - 2)$ circumnuclear or orbital electrons. The radioelement changes into an element two steps farther down in the procession of the elements.

[6] *Introduction*, p. 400.

This is the first half of the displacement-law of Fajans and Soddy. It signifies that *the emission of α-particles by a radioelement is the sign of a transmutation of that element into the next but one of those preceding it in the procession of the elements.* If the properties of this latter element are known already, the law can be tested with all the accuracy desired. Polonium stands two places after lead in the procession; it emits alpha-particles; its atoms should turn into atoms possessing all the chemical qualities of lead, and they do. If a radioelement which lies two steps ahead of an element not previously known is discovered to emit alpha-particles, we are still not without information as to the qualities of the element into which it should transmute itself. For if we know the column of the Periodic Table in which the original element lies, we know also the column in which the element two steps ahead of it should lie; and the chemists know what features are common to all the known elements of that column and presumptively extend also to the unknown one. Radium lies in the second column of the Periodic Table; it emits alpha-particles; it should be transmuted into an element lying in the "zero" column. That element was not known until after radium was discovered; but it was known that the other elements of the zero column are inert gases, and consequently that the one into which radium transmutes itself should be an inert gas. This is verified; and as a general rule it is verified that when a radioelement emits alpha-particles the substance left behind possesses the particular chemical features of the elements belonging to that column of the Periodic Table to which the element two steps preceding the original one belongs. From this fact of experience it is only a short step to the first part of the Fajans-Soddy displacement-law—and a step which is put quite beyond criticism by the relations presently to be cited which connect the atomic weights of the radioelements.

The second half of the law relates to the other radioelements, those which eject electrons from their nuclei. When an atom of atomic number Z emits an electron from its nucleus, the nuclear charge increases to $(Z + 1)e$, which is sufficient to hold another electron beyond the Z electrons of the original family. The atom does pick up another electron which enters into the circumnuclear set (*not* into the nucleus); and it becomes an atom of nuclear charge $(Z + 1)e$ and $(Z + 1)$ orbital electrons. *The radioelement changes over into another which is one step farther up the procession of the elements.* This is the second half of the displacement-law of Fajans and Soddy.

The evidence for this second part is extensive; but on the whole it is not so imposing as the evidence for the first part. Largely this is

due to the difference between the two types of emission. Alpha-particle emission is violent and unique; positively-charged particles moving with a speed like theirs are not produced in any other way known to man. Beta-particle emission is considerably less violent, and there are so many known processes for producing fast-flying electrons that one must always keep in mind the possibility that some of the electrons proceeding from radioelements may be due to one or another of these; in fact, many certainly are. Perhaps the best way to state the evidence is this: every radioelement which does not emit alpha-particles transmutes itself into an element lying one step farther up the procession,[7] and all but one (actinium) of these elements is known to emit electrons, all of which agrees with the assumption that in each of these transmutations one electron is extruded from each participating nucleus. Stated thus, it may not sound very convincing; but if the second part of the Fajans-Soddy law were not true, we should hardly have failed thus far to find something definitely inconsistent with it.

Were gamma-rays without an accompanying beta-particle or alpha-particle to be emitted from a nucleus we could scarcely call the result a transmutation, since it would not affect the nuclear charge nor the electron-family of the atom. There is no reason for denying that this might happen; but I am not aware that it is known ever to happen, except in cases of nuclei which have just previously undergone a transmutation—cases which we shall eventually examine.

If now each radioelement is passing over into another element, one step before it or two steps behind it in the procession according as it emits beta-rays or alpha-rays—then it must be possible to draw up *genealogies* of radioelements, series of elements of which each member is transmuted out of the foregoing and transmutes itself into the following one. All of the known radioelements fall into one or another of several such series. To represent all these relations, and one more, it is convenient and suitable to draw a graph in which the atomic numbers of the elements are laid off horizontally, and their atomic weights are laid off vertically. Each element is represented by a point upon this graph; when the element transmutes itself it moves to another point, two units to the left if an alpha-particle is emitted and one to the right if the change is a beta-ray change. Now the emission of an alpha-particle involves the departure of four units of mass from the nucleus which it leaves; the loss of an electron however involves a loss

[7] More precisely, into an element having the chemical features distinguishing the column of the Periodic Table containing the element lying one step farther up the procession than the original one.

of only 1/1850 of a unit of mass, which is quite inappreciable.[8] Hence in a transmutation of the former sort, the point representing the element in the graph moves four units downward as well as two to the left; in one of the latter sort, the point simply slides horizontally to the right. The meaning of Fig. 1 will now be clear.

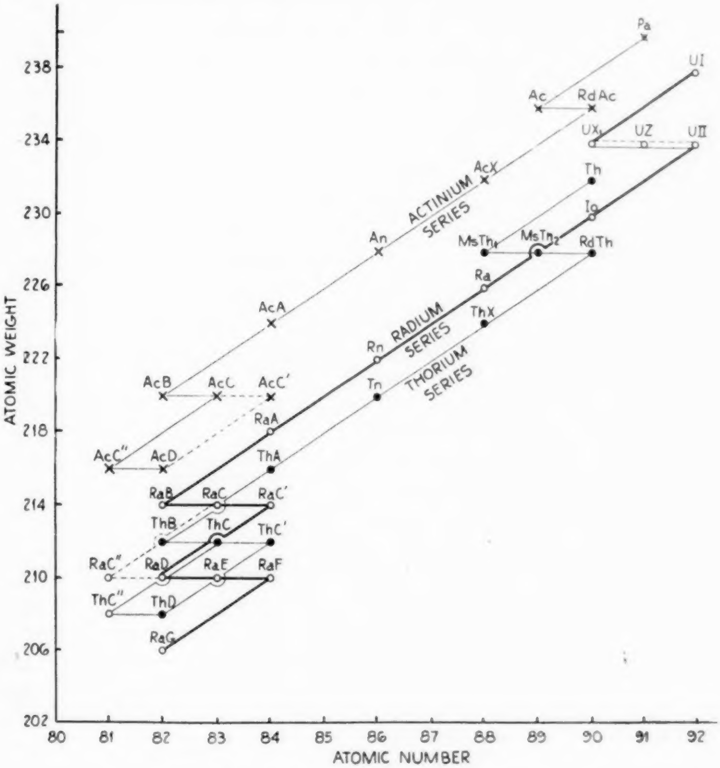


Fig. 1. Genealogies of the radioelements

(The actinium series is plotted some distance above the others for legibility, but should almost certainly lie lower.)

The lines in Fig. 1 which represent the family trees of the radioelements descend in zigzags, which signifies that the "decline and fall"

[8] These masses are given in terms of the unit of mass in which atomic weights are measured, of which 16 constitute the mass of an oxygen atom. Were the mass of an electron appreciable in these measurements, we should have to allow for the electrons added to or lost from the circumnuclear family to balance the change in the nucleus-charge. But then we should also have to make decisions about the mass to be assigned to the energy carried away by the particles and the waves.

of a radioactive atom does not proceed continually downward to lower and ever lower atomic numbers, but is interrupted by occasional partial recoveries. Whenever there are three consecutive transmutations of which two involve the emission of beta-rays and one that of alpha-rays, the element in which the third ends has the same atomic number as the element from which the first originates. In each of the three lines of descent made visible in Fig. 1 there are instances of this; the one including radium for instance touches three times at atomic number 82, the one commencing with thorium twice. In the sense in which I have thus far used the word "element," the element 84 recurs three times in one series and twice in the other. Here is an ambiguity which the time has come to dispel.

The ambiguity in the use of the term *element* is a question of words, but not wholly a linguistic, much less a trivial one; it is such a question as arises when a field of knowledge is expanded and enriched to such an extent that its old vocabulary ceases to be adequate. This particular question arose after the discovery that certain substances differing in radioactivity are very much alike in their chemical properties—another of the facts which the atom-model is especially adapted to explain.

Consider, for an example, the three elements radium B and radium D and radium G, which lie upon the same line of descent, the "radium series." The first transmutes itself into the second, and the second transmutes itself into the third, each in a three-stage process involving the departures of two electrons and an alpha-particle (with the order of their exits we are not now concerned) from the nucleus. The mass of the third is four units less than that of the second and eight units less than that of the first; in radioactivity also they differ. But all three are alike in nuclear charge, and hence in the size and presumably in the arrangement of their circumnuclear families of electrons; and hence the presumption arises, that in their physical and chemical properties apart from radioactivity and mass they should be quite alike.

This presumption about the chemical properties is confirmed by the fact that RaB and RaD and RaG cannot be separated from one another by chemical means after they are once mixed. In general, whenever two of these "radioelements" coinciding in atomic number are subjected to any of the very considerable variety of agents in the chemists' armory, they respond in so nearly, if not exactly, the same way that there is no method known for taking one and leaving the other. Crystallization out of a mixture of salts of two such elements merely produces crystals containing the two salts in the same proportion as the liquid; sublimation merely produces a deposit containing the elements

in their original ratio; electrolysis does not favor one above the other, nor does osmosis, and if a mixture of the two elements or of salts of theirs is presented to an absorbent or an adsorbent or a solvent, it accepts them in the same proportions as they are presented, while any element willing to react with either reacts in precisely the same degree with the other.[9] Similarity such as this goes far beyond the interresemblances of the alkali metals, for instance, or even those of the rare-earth elements, difficult as the task of separating these latter from one another sometimes proves; it is not similarity merely, it amounts to identity.

As for the presumption that radioelements sharing the same atomic number should be alike in what are loosely termed the "physical properties," it is more difficult to test. In fact, there seems to be no instance of two such elements, each radioactive and each obtainable quite unmixed with the other in quantities large enough for such experiments. The three elements sharing atomic number 86 are all gaseous at ordinary temperatures, but they are too scanty and two of the three are much too fugitive for making accurate comparative measurements of such qualities as viscosity or elasticity or ionizing potential. The elements sharing atomic number 82 are, as I shall presently bring out, mostly stable, and upon them it is possible to test the expected coincidence in optical line-spectra and X-ray spectra, which is verified except for certain very minute (but unexplained!) differences in the wavelengths of certain lines. The band-spectra of these elements (more precisely, of their compounds) display slight differences which are beautifully explained by the contemporary theory of band-spectra, involving as it does a participation of the nucleus with its mass in the production of the bands.[10] Mixtures of two of the elements sharing atomic number 90 (thorium and ionium) display precisely the same optical spectrum as pure thorium.[11] In addition to these observations, a great many have been made upon the physical

[9] There is a huge literature of the attempts to separate elements of identical atomic number and to discriminate between their chemical properties, for a review and bibliography of which I refer again to von Hevesy and Paneth (*l.c. supra*, chapter XII).

[10] E. S. Bieler, *Nature*, **115**, p. 980 (1925).

[11] This is vividly illustrated by a passage in the classical treatise upon radioactivity which Rutherford wrote in 1912. Boltwood had isolated from uranium ores a sample of thorium oxide which emitted, along with the alpha-particles from the thorium, a considerable number coming from ionium. Russell and Rossi produced its arc spectrum and "the spectrum of thorium was obtained, but not a single line was observed that could be attributed to ionium. On the assumption that ionium has a life of 100,000 years, the preparation should have contained 10% of ionium. Since probably the presence of 1% of ionium would have been detected spectroscopically, it would appear that the ionium was present in small amount, indicating that the life of ionium must be much less than 100,000 years." As a matter of fact there was probably more than 10% of ionium in the mixture; but its spectrum lines were identical with those of thorium.

properties of other non-radioactive elements which share particular atomic numbers and are mixed together in varying proportions; and they establish that such "elements" are indistinguishable except in such properties as are influenced to a measurable extent by the mass of the nucleus.

These facts make it necessary to redefine the word *element*, which in its long journey through the centuries from Lucretius has modified its meaning time and time again to keep pace with the gradual refinement of scientific thought, though all the while it kept its spelling intact. These are the alternatives: *either* to confer the status of a separate "element" upon each substance (apart, of course, from the compounds!) possessing a distinctive mass and radioactivity of its own, so that there may be several distinct elements sharing a given set of chemical properties—*or* to link the term "element" with a characteristic set of chemical and physical properties, with a specific atomic number and position in the Periodic Table, so that a given element may be an ensemble of several different kinds of matter differing in radioactivity or mass or both. Reasons of science require that one or the other of these alternatives be chosen, but the actual choice is determined by reasons of language and expediency. These reasons—I will not pause to develop them—favor the second alternative. Inconvenient though it may be to refer to RaB and RaD and RaG and ThB and several other radioactive substances as the same element, the inconveniences entailed by the other policy would in the end be immensely greater. One element to each atomic number, one place in the Periodic Table to each element—this is the choice which the prior usage and the associations of the word *element* recommend; and some other name must be selected to distinguish the substances which share a common atomic number but differ in mass or radioactivity or both.

Such a name is Soddy's word *isotope*, constructed out of Greek words to signify "in the same place." Radium B and RaD and all the other substances which appear in the column labelled "82" in Fig. 1 are isotopes of the element 82; radon and thoron and actinon are isotopes of the element 86. In these eleven places of the Periodic Table extending from 81 to 92, the individual isotopes enjoy names of their own, the elements are best known by their numbers. The names *thallium, lead, bismuth* and *uranium* are, it is true, generally attached to the elements 81, 82, 83 and 92; but the first three of these names are used by some people to mean the elements in question and by others to designate only those of their isotopes which are not radioactive, and there is danger of confusion.[12] Elsewhere in the Periodic Table, where

[12] The names *polonium, radium, actinium, thorium* and *protactinium* signify par-

all the isotopes of each element are stable, the elements have individual names and the isotopes are designated only by their masses. The elements 81, 82 and 83 have some isotopes which are radioactive and others which are not; thus the word "radioelement" is misleading, and should be replaced by "radioactive isotope." Consistency indeed requires that one speak of the successive members of a family of radioactive substances not as consecutive elements, but as consecutive isotopes of diverse elements. At this point however consistency almost ceases to be a jewel. I can find no satisfactory compromise, and will hereafter refer to the various radioactive materials simply as "substances"—so bringing to an end this long analysis of words, which is justified only in so far as it may have concentrated the reader's attention upon the facts underlying them.

We return to Fig. 1.

The radioactive substances are grouped into three main lines of descent or sequences, commonly called *series*. Each of these throws off one or two branches, which however cannot be followed far; these I will discuss further on, pausing here only to mention that one of the three main sequences, the *actinium series*, is believed by many to branch in this manner out of the *uranium-radium series*. This however is not certainly established, and it is suitable to regard these two and the *thorium series* as independent sequences, which between them comprise all the known radioactive isotopes among the elements.[13]

Uranium and thorium, the first elements of the series to which they have given their names, are even yet after all the aeons of the earth's existence to be found in abundance among its rocks. This practically proves that uranium, at least, disintegrates with exceeding slowness; for all the other known elements are lighter than it is, and consequently there is none of them out of which the steadily-dwindling supply of uranium might be replenished by transmutation. We shall presently learn methods of estimating the duration of uranium, by which it is shown to be truly colossal.

The atomic weights of uranium and thorium are known, and amount to 238.18 and 232.12 respectively. From these it should be possible

ticular isotopes of the elements 84, 88, 89, 90 and 91 respectively, but are sometimes used as names for these elements—another dangerous source of misunderstanding. The name *niton* was formerly used for the isotope *radon* of element 86, and might well be used for this element now that the isotopes are individually named.

[13] Apart from the elements potassium and rubidium, which will continually demand to be mentioned as exceptions unless they are disposed of once for all at this point. Let it be stated, then, that these elements emit electrons, so feebly however that they are much less active than even uranium, which ranks among the least radioactive of all the known radioactive substances; and that no one has identified the substances into which they are transmuted, though presumably those are isotopes of calcium and strontium respectively. Cf. an account of the radioactivity of these elements by A. Holmes and R. W. Lawson: *Phil. Mag.* (7) **2**, pp. 1218–1233 (1926).

to deduce the atomic weights of all the other members of the two sequences; thus, a radium atom is what is left behind after a uranium atom has ejected three alpha-particles (mass, 4 apiece) and two electrons (mass negligible) and its atomic weight should therefore be 226.18. Here we meet a troublous fact. The value of the atomic weight of radium, as measured by no less an expert than the celebrated Hönigschmid, is 225.97 with an uncertainty believed not to exceed three units in the last place. This might be explained by supposing that the *element* uranium as found in nature is a mixture of several isotopes in relatively large proportions, only one of which is the parent of the uranium-radium series, while the others may be stable or perchance the ancestors of the other series; indeed it is hard to think of any other adequate explanation.[14]

All three of the sequences terminate in isotopes of the element 82, commonly known (but remember the caution on page 110!) as lead. It is a curious fact that the most rare and precious of all substances should die away by self-transmutation into the one which serves as the symbol for everything which is commonplace, dull and cheap. The atomic weights of the terminating isotopes of the radium and thorium sequences may be guessed in the same manner as that of radium from that of uranium. Starting from radium and from thorium respectively and noting that an atom of radium is destined to eject five alpha-particles and an atom of thorium six during the transformations whereby they turn into atoms of RaG and ThD respectively, we calculate the values 206.0 and 208.1 for the atomic weights of these two isotopes of element 82. Now nearly every sample of lead that has ever served for an atomic-weight determination has yielded a value near 207.2. Yet, when the lead-content of certain minerals rich in uranium and its posterity and deficient in thorium was extracted and investigated, the atomic weights of these samples were found to lie extremely near to 206—some of the values recorded are 206.046, 206.048 and 206.08. On the other hand, samples of lead extracted from various minerals rich in thorium and poor in uranium displayed abnormally high atomic weights, values attaining in some instances to 207.9. These are data much more dramatic than the customary outcome of the tedious process of determining an atomic weight; one wonders vainly what chemists would have felt, if they had been published before

[14] What is commonly called "uranium" contains not only the ancestor of the uranium-radium series, but also one of its descendants, which however is not present in sufficient amount to affect the atomic weight. This is the reason for inserting the words "in relatively large proportions" in the above sentence. The fact that the atomic weight of uranium is not integral might be taken to suggest that it is a mixture of integral-weight isotopes. Aston's latest experiments on stable elements of non-integral atomic weight show, however, that the premise does not necessarily lead to the conclusion.

radioactivity was discovered. They disclose the only known instance of distinct stable isotopes of an element being found separately from one another in nature. Whether "ordinary lead" of atomic weight 207.2 is a mixture of these two isotopes, or contains still others, is as yet an unsolved question.[15]

The three series resemble one another not only in the nature of their terminal substances, but in other regards as well. The substance in the radium series known as ionium, the member of the actinium series called radioactinium, the member of the thorium series named radiothorium, are isotopes all three of the element 90; and these three substances evolve through the same succession of transformations, alpha-ray emissions and beta-ray emissions following after one another in the same order. The *n*th descendants of these three substances, for each value of *n* from 1 to 6, are isotopic with one another—a statement which will probably be made clearer by Fig. 1 than by these words. This parallelism, which from the grandchildren of Io and RdAc and RdTh onward is reflected in the names of the substances, includes also the "branchings" which occur in each sequence at the substance labelled C—radium C and actinium C and thorium C. It is limited in its range, for the earlier parts of the three sequences are by no means alike, while the radium sequence continues onward for three stages longer than the two others. Something within the radium atom impels it to continue evolving even after it has twice taken and left the atomic number which it is destined eventually to take and keep, although the atoms which were once actinium or thorium are contented to stop at the atomic number 82 when for the second time they reach it.

The phenomenon of *branching*, which I have twice casually mentioned, is worthy of a few paragraphs. It signifies that a certain proportion of the atoms of such a substance as (for instance) thorium C transmute themselves in one fashion, the remainder in another. Sixty-five per cent of the atoms of ThC extant at any moment are destined to emit beta-rays and become atoms of a substance ThC′ lying one step further up the procession of the elements; the other thirty-five per cent eventually emit alpha-particles and become atoms of ThC″ placed two steps further down the procession. Such a "dual transmutation" occurs also at RaC and at AcC—an instance of the parallelism just mentioned, which however does not extend to the relative frequency of the two modes of transformation; 9996 out of ten thousand atoms of RaC, but only three out of a thousand atoms of

[15] Not however a definitely insoluble question, since the Thomson-Aston method of resolving mixtures of isotopes (*Introduction*, pp. 14–29) and measuring their individual masses should be applicable to lead—that is to say, certain difficulties have thus far prevented it from being applied to the very heavy elements, but these difficulties may not prove insuperable.

AcC, transmute themselves by ejecting electrons. As the disintegration of a sample of any of these substances proceeds, the relative proportions of the atoms disintegrating in the two ways remain unchanging. This makes it seem inadvisable to describe ThC (for instance) as a mixture of two distinct substances; rather it appears that the atoms may be all alike, but the destiny of each particular atom is a matter of "chance," with the chances favoring one type of disintegration over the other by nearly two to one. This is not the only circumstance in radioactivity which suggests the operations of "chance."

The substances labelled C′ and C″, which result from the dual disintegration of any of the three substances labelled C, differ in atomic weight and in atomic number, and in radioactivity as well; for the C′ substances which were born out of beta-ray transformations emit alpha-rays, while the C″ substances which resulted from alpha-ray transmutations send forth beta-rays. Consequently their immediate descendants, the two grandchildren of each C-substance, are isotopes with one another—and isotopes which should be alike not only in atomic number but in atomic weight as well. Is there any respect in which they differ? We cannot tell. Both of the grandchildren of ThC are apparently non-radioactive and stable; probably they are one and the same isotope of lead. Both grandchildren of AcC likewise seem to be stable. The predominant grandchild of RaC is the radioactive substance RaD; but in this case the number of atoms of RaC electing the less popular path of disintegration is so exceedingly small that we can neither discern any distinctive radiation to be ascribed to a substance isotopic with RaD but distinct from it, nor yet conclude from our failure that no such substance exists. Concerning the fourth of the known branchings, which occurs at UX_1, the state of affairs is the same as with RaC; we can neither detect more than one kind of grandchild, nor be sure that there is only one. In this case, by the way, both modes of transmutation of the parent element involve the emission of beta-rays.

Although among the four substances which are known to disintegrate in two alternative ways there is thus none for which both of the two lines of posterity can be traced through more than two generations, it is believed by many that there must be a fifth such substance in the uranium series, from which the actinium series goes off as a branch while the main proportion of the atoms continue evolving down the radium sequence. The reason for this idea is that in the ores of uranium the members of the actinium sequence are as a rule to be found about three per cent as abundantly as the members of the radium sequence. This fact could be deduced by assuming that

uranium II suffers a dual alpha-ray disintegration, about 97 per cent of the atoms transmuting themselves into ionium and the other 3 per cent into the mysterious substance uranium Y which is always found mixed with uranium, and which is known to emit beta-rays and hence to pass over into an isotope of element 91 which may well be protactinium, the first known member of the actinium series. On following out these presumptive transformations in Fig. 1 the reader will see that they would lead to the actually-observed result; but that is not quite the same thing as proving that the observed result is attained in just that way. The branching may occur elsewhere in the posterity of uranium; or the observed constancy of the ratio of actinium to radium in the rocks may mean that actinium and its family all descend from a separate isotope of element 92, not concerned in the production of radium. Much light would be shed upon this question if someone would only determine the atomic weight of even one member of the actinium sequence—an achievement which would settle at once those of all the others, and is most eagerly awaited.

Having dealt with the filiation of the radioactive substances, having specified the substance from which each is born and the substance to which it gives birth, and the sort of particle which is emitted in each process of transmutation, it remains to specify the rates at which the transmutations occur, and the speeds of the particles which are emitted, and the wavelengths of the quanta of radiation which sometimes come out also, and how many there are of these. The fundamental assumptions of the theory of radioactivity, which the experiments have sustained, require that in a transmutation only one alpha-particle or one beta-ray be emitted from the nucleus of one self-transmuting atom; but there is no such limitation upon the radiation-quanta, nor upon the electrons incidentally ejected from the circumnuclear family.

The rate of transmutation of every radioactive substance, so far as we know, is governed by the famous exponential law which signifies that *equal fractions perish in equal times*—that if one were to take a sample of the substance and determine the quantities extant at two instants an hour apart, and also those existing at two other instants an hour apart, and at any number of pairs of instants separated by intervals of one hour, then the mutual ratios of the two measured values of all those pairs would be the same. Half of any sample of thorium C transmutes itself in one hour; half the remainder in the next hour; half the remainder in the next hour, and so forth *ad infinitum* (or, to speak more carefully, up to the limit of the observations).

This law is described by the following formula relating the quantity Q of the substance existing at any time t, and the quantity Q_0 existing

at any other time t_0 (provided that no replenishment of the supply is taking place!):

$$Q = Q_0 \exp\left(\frac{t_0 - t}{\tau}\right). \tag{1}$$

Furthermore the rate dQ/dt at which the substance is being transmuted at any instant is related to the amount Q existing at that instant as follows:

$$dQ/dt = -\frac{Q}{\tau} = -\frac{Q_0}{\tau}\exp\left(\frac{t_0 - t}{\tau}\right). \tag{2}$$

These formulæ contain only a single constant characteristic of the substance. Nothing simpler could be desired. A phenomenon that

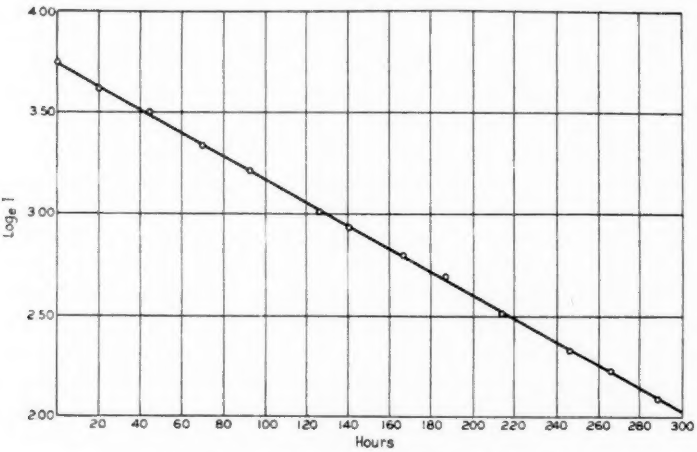


Fig. 2. Decay-curve of radium E (R. F. Curtiss)

(Being plotted on logarithmic paper, the graph of the exponential law is a straight line.)

can be described by a formula involving only one constant which has to be varied to distinguish one case from another is a rare gift of nature.

While the equations (1) and (2) are naturally valid whatever the unit in which we choose to measure Q, it is desirable as a rule (and necessary, in comparing the radioactivity of different substances) to express Q either in gramme-molecules, or in actual numbers of atoms. In some places I shall use N as a symbol for Q measured in the latter manner.

The exponential law is a law of chance. It may be expressed by saying that the chance of an atom disintegrating within a given time-

interval is precisely the same, whichever atom one chooses to consider and whenever the instant at which one chooses to let the given time-interval begin. I will quote a passage from Poincaré [16], taking only the liberty of writing 'nucleus' where he wrote 'atom.' "If we reflect on the form of the exponential law, we see that it is a statistical law; we recognize the imprint of chance. In this case of radioactivity, the influence of chance is not due to haphazard encounters between atoms or other haphazard external agencies. The causes of the transmutation, I mean the immediate cause as well as the underlying one (*la cause occasionnelle aussi bien que la cause profonde*) are to be found in the interior of the atom [*read*, in the nucleus]; for otherwise, external circumstances would affect the value of the coefficient in the exponent. . . . The chance which governs these transmutations is therefore internal; that is to say, the nucleus of the radioactive substance is a world, and a world subject to chance. But, take note! to say 'chance' is the same as to say 'large numbers'—a world built of a small number of parts will obey laws which are more or less complicated, but not statistical. Hence the nucleus must be a complicated world. . . ." I shall make no further allusion to theories of radioactivity.[17]

The constant τ may be interpreted as the time-interval during which the fraction $\frac{e-1}{e}Q$ (or approximately $0.632Q$) of any initially-present quantity Q of the substance would undergo its change. It is greater by the factor $1/\log_e 2$ (or approximately 1.44) than the *half-period* of the substance, the interval (designated by T) during which one half of the initially extant atoms are transmuted. It is also the average duration of the life of a single atom. All of these statements may be proved without difficulty from the formula (1). From the similarity between (1) and (2) it follows that the rate at which transmutations occur in an unreplenished sample of a radioactive substance, and the rate at which rays shoot out of such a sample, and the intensity of all the effects which the rays produce, vary exponentially with time; and the constant T which is the half-period for the extant quantity of the substance is likewise the half-period for all of these. The constant τ likewise has the same meaning for them all, and so does its reciprocal

[16] *Dernières Pensées*, pp. 204–205; he credits Debierne with the idea.

[17] Further and very valuable evidence that the transmutations of individual atoms are governed by the "laws of chance" operating within their own nuclei is furnished by the variations or fluctuations (*Schwankungen*) of the numbers of alpha-particles emitted from a sample of any radioactive substance in consecutive equal time-intervals very short compared with the half-period of the substance (*v.i.*). These are precisely analogous to the fluctuations in thermionic emission known by the name of "Schroteffekt" (***Introduction***, p. 10). Consult an article by K. W. F. Kohlrausch in ***Ergebnisse der exakten Naturwissenschaften***, **5** (1926).

λ, which is called the *disintegration constant*, and is often specified instead of τ or T.

The values of the half-period T for the various radioactive substances are collated in the accompanying Table, which contains also the names of the substances, their usual symbols (those used in Fig.1), the symbols embodying their atomic numbers proposed by Kovarik and McKeehan, and the types of particle which they emit from their nuclei.

SYMBOLS, NAMES AND HALF-PERIODS OF THE RADIOACTIVE SUBSTANCES

The first column of this Table contains the usual symbols for the substances; the third, their usual names; the fourth, their half-periods as collated by A. F. Kovarik and L. W. McKeehan (*l. c. supra*); the fifth, the nature of the particles which they emit at transmutation. In the second column, the symbols proposed by Kovarik and McKeehan are given; each is composed of the atomic number of the substance, of a symbol denoting the series to which it belongs, and sometimes of a second numeral which, when the substance is an isotope of one or more others in the same series, denotes whether it is the first, second or third of these isotopes reached in the course of the transmutations. In cases of branching, the less common of the two resulting substances is italicized. The annotation *est.* signifies that the half-periods in question are estimated by extrapolating the Geiger-Nuttall relation (*v.i.*). The abbreviations *s*, *m*, *d*, *a* stand for second, minute, day, year.

URANIUM-RADIUM SERIES:				
UI	92UI	Uranium I	$4.6 \cdot 10^9 a$	α
UX_1	90UI	Uranium X_1	$24.5d$	β to UX_2, β to UZ
UX_2	91U	Uranium X_2	$1.138m$	β
UZ	*91Ua*	*Uranium Z*	*6.69h*	β
UII	92UII	Uranium II	$1.2 \cdot 10^8 a$ (*est.*)	α
Io	90Ra	Ionium	$7.43 \cdot 10^4 a$	α
Ra	88Ra	Radium	$1.69 \cdot 10^3 a$	α
Rn	86Ra	Radon, radium emanation	$3.810d$	α
RaA	84RaI	Radium A	$3.0m$	α
RaB	82RaI	Radium B	$26.8m$	β
RaC	83RaI	Radium C	$19.5m$	β to RaC′, α to RaC″
RaC′	84RaII	Radium C′	$10^{-6}s$	α
RaC″	*81Ra*	*Radium C″*	*1.32m*	β
RaD	82RaII	Radium D	$16a$	β
RaE	83RaII	Radium E	$4.85d$	β
RaF	84RaIII	Radium F, polonium	$136.3d$	α
RaG	82RaIII	Radium G, radium lead apparently stable		
THORIUM SERIES:				
Th	90ThI	Thorium	$1.3 \cdot 10^{10} a$	α
MsTh1	88ThI	Mesothorium 1	$6.7a$	β
MsTh2	89Th	Mesothorium 2	$6.20h$	β
RdTh	90ThII	Radiothorium	$1.90a$	α
ThX	88ThII	Thorium X	$3.64d$	α

Tn	86Th	Thoron, thorium emanation	54.5*s*	α
ThA	84ThI	Thorium A	0.145*s*	α
ThB	82ThI	Thorium B	10.6*h*	β
ThC	83ThI	Thorium C	60.6*m*	β to ThC′, α to ThC″
ThC′	84ThII	Thorium C′	10^{-11}*s* (*est.*)	α
ThC″	*81Th*	*Thorium C″*	*3.20m*	β
ThD	82ThII	Thorium D, thorium lead apparently stable		

ACTINIUM SERIES:

Pa	91Ac	protactinium	$1.6\cdot10^{4}$*a*	α
Ac	89Ac	actinium	20*a*	β
RdAc	90Ac	radioactinium	18.9*d*	α
AcX	88Ac	actinium X	11.2*d*	α
An	86Ac	actinon, actinium emanation	3.92*s*	α
AcA	84AcI	actinium A	2.00*s*	α
AcB	82AcI	actinium B	36.1*m*	β
AcC	83Ac	actinium C	2.16*m*	α to AcC″, β to AcC′
AcC′	*84AcII*	*actinium C′*	*.009s*	α
AcC″	81Ac	actinium C″	4.71*m*	β
AcD	82AcII	actinium D, actinium lead apparently stable		

UY, K, Rb not assigned to series. They emit beta-rays, and their half-periods are given respectively as 24.6*h* (St. Meyer, *l. c.* footnote 1), $1.5\cdot10^{+12}$*a* and 10^{11}*a* (Holmes and Lawson, *l. c.* footnote 13).

To measure a disintegration-constant seems an easy task, since one has only to choose the most convenient effect of the rays of the substance in question, and measure it at sufficiently many times to establish a sufficiently long arc of its decay-curve. Yet there is, I suppose, no other problem of which the general solution involves as many of the typical difficulties of research in this field; partly because some of the half-periods to be measured are so exceedingly short and some so tremendously long, largely because no radioactive substance ever exists by itself. Some can be separated completely from their ancestors, but none can ever be totally isolated from its posterity, especially since its rate of producing its posterity is the very thing which is being measured. Its own gradually-declining rays are mixed with the gradually-augmenting rays of its descendants, and while the specific effects of the former can indeed in some cases be distinguished from those of the latter, this is often difficult and sometimes impracticable. Frequently the observer is required to deduce the half-periods of individual substances from observations upon a continually-changing mixture; and most of the mathematical formulæ used in the study of radioactivity are developed out of equations (1) and (2) for interpreting such observations, or inversely for predicting the evolution

of a mixture of substances of which the initial composition is taken for granted. There is no better way of conveying a notion of the methods by which radioactivity was and is studied than to describe how some of the known half-periods were actually ascertained.

The simplest of all the cases are those in which a substance which can easily be separated from its ancestors transmutes itself into one which either is not radioactive at all, or else decays so slowly that the rays which it emits are not strong enough to interfere with the observations on the rays of its parent. The penultimate substances of the various series are candidates for this class, but the only one among them which is abundant enough and lasts long enough to be easily isolated from its ancestors is radium F, otherwise known as polonium. This therefore is the classical instance of a substance of which the decay-curve is determined directly from observations on rays of its own. Another is radium E, of which the half-period is so short (about 5 days) and the half-period of its daughter-substance so long (more than four months) that its decay-curve can be traced practically as if it changed into a stable element (Fig. 2).

Almost as simple are certain cases in which a radioactive substance is isolated both from its ancestors and from its posterity, and then the growth of its immediate descendant is measured. This method is available when the parent-substance is much longer-lived than its child, so that the rate at which atoms of the latter come into being is practically constant throughout the period of observation. Let B represent this rate; let Q represent the quantity of the daughter-substance extant at any time t, the time being measured from the instant when the isolation of the parent-substance is perfected, so that $Q = 0$ at $t = 0$; let λ stand for the disintegration-constant of the daughter-substance, so that the rate at which its atoms are disappearing through transmutation is equal to λQ. The net rate of growth of the daughter-substance is therefore

$$dQ/dt = -\lambda Q + B \tag{3}$$

from which we obtain by integration

$$Q = \frac{B}{\lambda}(1 - e^{-\lambda t}), \tag{4}$$

so that the quantity of the daughter-substance, and the intensity of its rays vary as exponential functions of time with the disintegration-constant standing in the exponent. This function, it is true, rises from zero to a positive final limiting-value instead of falling to zero from a positive initial value, as the decay-curve would; but the value

of λ is determined from it quite easily, and as a matter of fact the decay-curve itself can be obtained merely by plotting as function of time the difference between Q and the limiting-value ($= B/\lambda$) which Q approaches as t increases indefinitely. Determining a half-period from a rate of growth is therefore mathematically the same process as determining it from a rate of decay. This is one of the ways in which the half-period of uranium X_1 is measured; and the standard method for determining that of radium is based partly upon it, as we shall presently see.

Eventually the grandchild and the remoter posterity of the parent-substance must make their presence known. This is not always a disadvantage. Letting λ_1 and Q_1 stand for the disintegration-constant and the extant quantity of the daughter-substance, λ_2 and Q_2 for those of the granddaughter, we have as basis for the theory these equations:

$$dQ_1/dt = B - \lambda_1 Q_1, \qquad dQ_2/dt = \lambda_1 Q_1 - \lambda_2 Q_2, \tag{5}$$

integrating which, and supposing that at $t = 0$ the parent-substance has just been isolated so that the building-up of the two descendants from zero is just commencing, we obtain for Q_1 the expression (4) with λ_1 in the place of λ, and for Q_2 the function

$$Q_2 = B\left[\frac{1}{\lambda_2} + \frac{1}{\lambda_1 - \lambda_2}e^{-\lambda_1 t} - \frac{\lambda_1}{\lambda_2(\lambda_1 - \lambda_2)}e^{-\lambda_2 t}\right], \tag{6}$$

which to second approximation is equivalent to

$$Q_2 = \frac{1}{2}B\lambda_1 t^2. \tag{7}$$

The amount of the grandchild therefore should increase at first as the square of the time elapsed, whereas the amount of the child increases proportionally to the time. There are instances, in the history of the study of radioactivity, of a substance being regarded as the child of another until measurements were made upon its rate of growth in an isolated sample of its putative parent, whereupon through its conformity to (7) it was proved to be the grandchild and not the child. The question whether radium comes directly out of uranium II, or out of an intermediate substance, was settled in this fashion; and by observing a sample of uranium II at intervals over a period of almost twenty years, and measuring the radium which was being developed within it, Soddy was able through equation (7) to calculate the half-period of this intermediate substance (ionium).

The method used in deriving the equations (4) and (7) can always be

applied to any number of consecutive radioactive substances; there are always just equations enough to determine all the constants and describe completely the future history of any mixture of the members of a single family line, provided that their relative proportions in the mixture are specified for some particular moment. Even with only three substances the behavior of the mixture may be extraordinarily complicated; but there are simpler cases which are instructive.

If for instance one sets aside a substance with a much longer half-period than any of its posterity possesses, the extant quantity of each and every one of the descendants will first increase and then begin to decrease, and eventually diminish along the same exponential curve as the long-lived ancestor itself—not because the half-periods of the descendants are actually changed, but because of the partial balancing between the decay and the replenishment of each. Thus the half-period of the long-lived ancestor may be determined by plotting against time the total intensity of all its rays and all the rays of its descendants, or that of any particularly convenient kind of ray emitted by any member of the family. The most carefully measured and accurately known of all disintegration-constants, that of radon, is usually determined in this way; its half-period amounts to four days, those of its three next descendants radium A and radium B and radium C to only a few minutes each, so that after isolating a sample of radon and waiting a few hours one can set up any device for measuring the gamma-rays of radium C, plot their decay-curve, and from it determine a value of λ which is not that of radium C, but that of radon.

If in such a case as the foregoing the long-lived ancestor is so very long-lived that no appreciable decrease in its rate of transmutation can be detected over a period of years, then eventually the quantities and the radiations of all of its descendants assume values which likewise do not change appreciably for years; "radioactive equilibrium" is attained. In a unit of time, equal numbers of atoms are transmuted out of each substance into the substance following, into each substance out of the one preceding. Representing by N_n the number of atoms of the nth member of the series (counting the very long-lived ancestor as the first) extant in the mixture in radioactive equilibrium, by λ_n its disintegration-constant, and remembering that $\lambda_n N_n$ is the rate at which its atoms perish by transmutation, we have the chain of equations:

$$- dN_1/dt = \lambda_1 N_1 = \lambda_2 N_2 = \lambda_3 N_3 = \lambda_4 N_4 = \ . \ . \ . \qquad (8)$$

from which, if we know the relative quantities of any two members in

a mixture in equilibrium, and the half-period of either, we can determine the half-period of the other.[18]

This method could be applied to estimate the half-period of radium, which is so long that in the years since it was first isolated no sample has yet become perceptibly feebler in emitting its rays, while the half-periods of its descendants are all much shorter, and that of its child is only 3.82 days and is rather accurately known. However, the volume of radon gas in equilibrium with one gramme of radium (about the largest quantity of radium which has ever been gathered together in one place) is at normal temperature and pressure only about .0006 cc, and the measurement of so small a quantity of gas is inevitably so inexact that this method cannot compete even with the not-very-accurate alternative methods which we shall presently meet. However, its results do not disagree with theirs.

The most fascinating application of this method is made upon the rocks of the earth, which have presumably been existing so long that there has been ample time for the longest-lived member of the uranium-radium series to attain equilibrium with all of its descendants. As it happens, the longest-lived member of this series is the first, uranium I. Probably this is no mere accident; if uranium is the descendant of less lasting ancestors, they would all be gone by now. However that may be, it is a fair presumption that at least the older rocks of the earth have been formed and buried long enough for the uranium in them to have attained to equilibrium with its descendants. The ratio of the concentration of uranium to the concentration of any member of its posterity, radium for example, should then be equal to the reciprocal of the ratio of their half-periods. Great numbers of samples of rock from all over the world were analyzed by Rutherford and his pupils, and in the laboratories of France and Germany; and for a large proportion among them the ratios of the radium content to the uranium content were found to lie close to one another, and to a mean value which Rutherford assigns as $3.40 \cdot 10^{-7}$. Accepting this as the equilibrium-ratio, and 1690 years as the half-period of radium, we obtain for the half-period of uranium the truly colossal figure of 4.4 billions of years! This value is substantiated, as we shall presently see, by an altogether different method.[19]

[18] In some of the older rocks of the earth, uranium and its descendants have attained mutual equilibrium, and the value of λN for uranium in such a rock is equal to the rate at which the inert end-product (RaG) of the series is accumulating, so that by measuring the amount of RaG already accumulated and the amount of uranium still remaining one can estimate the age of the rock. Consult O. Hahn, *Handbuch der Physik*, **22**, pp. 289–306.

[19] This is a fortunate circumstance, as it gives greater confidence in rejecting the data obtained with samples of rock which yield values of the radium-to-uranium ratio differing considerably from $3.4 \cdot 10^{-7}$. In some cases these deviations may be

One of the two best methods for determining the half-period of radium is a combination of this last-named method with one of those which I described earlier. Let us suppose that a sample of ionium, equal to the amount which would be in equilibrium with one gramme of uranium and $3.40 \cdot 10^{-7}$ grammes of radium, is purified of its original radium-content and set aside for occasional observations of the rate of growth of fresh radium in it. Representing by N_1 the number of ionium atoms in the sample (which diminishes in so small a proportion that we may consider it constant), by N_2 the number of radium atoms extant at time t after the new supply begins to grow, by λ_1 and λ_2 the disintegration-constants of these two substances; translating equation (4) into this notation, and remembering that the rate of transmutation of the parent substance which was there called B is now (measured in atoms transmuted per second) equal to $\lambda_1 N_1$, we have

$$N_2 = \frac{\lambda_1 N_1}{\lambda_2}(1 - e^{-\lambda_2 t}). \tag{9}$$

Represent by N_{20} the number of atoms of radium which would be in radioactive equilibrium with the sample of ionium, that is to say, the number of atoms in $3.40 \cdot 10^{-7}$ grammes of radium; by equation (8) we have

$$\lambda_1 N_1 = \lambda_2 N_{20}, \tag{10}$$

so that equation (9) may be transformed into one containing no constants except the known one N_{20} and the object λ_2 of the investigation. The gain is still greater; developing the exponential function in (9) as a power-series in t and retaining only the first term, we have

$$N_2 = N_{20}(1 - e^{-\lambda_2 t}) = N_{20}\lambda_2 t + \text{terms of higher order}. \tag{11}$$

This means that we need to trace the growth-curve of radium out of ionium only so far as is necessary to determine its initial slope, the initial rate at which the radium increases before its own transmutation begins to tell. This as it happens is all that there has yet been time to trace, so that this combination of the two methods is the only way yet available of interpreting the growth-curves.[20] After a sample of ionium has been kept for a century or two, it may be possible to trace a long enough arc of the curve to determine by the first method. After

ascribed to the comparative youth of the rocks, in others to the selective action of flowing water and other geological agents in removing some and leaving others of the members of the radioactive family.

[20] This method, it will be perceived, is essentially a measurement of one and hence of all of the terms $\lambda_n N_n$ which are equated in equation (9); the rate of growth of radium out of ionium being ascertained, it is possible to calculate the value of λ_n for any substance in the radium series for which N_n, the quantity in equilibrium with the preassigned quantity of ionium, can be measured.

our descendants have solved the other problems of physics, they may be able to entertain themselves by keeping records of the behavior of long-lived radioactive substances, and so determining half-periods with an accuracy improving from millennium to millennium.

Another and the most picturesque of all the ways of determining a disintegration-constant consists in counting the atoms which in a measured quantity of the substance disintegrate in each second. It sounds almost unbelievable that this should be feasible; but it is really practicable to count the alpha-particles which proceed from a radioactive substance, for they make individual visible scintillations upon a fluorescent screen placed across their paths. If this device is inconvenient, one can measure the total charge which the alpha-particles carry into a chamber arranged to receive them, and divide it by the specific charge borne by each, which is very accurately known. The particles and consequently the transmuted atoms having been counted, it is necessary to weigh the substance which is emitting them; and this requirement is less easy to fulfil, being fulfillable in fact only for three substances—radium, and the long-lived ancestors thorium and uranium. Dividing the mass of the weighed sample by the mass of an atom, and dividing the quotient into the number of alpha-particles emitted per second, we obtain the value of λ_1. This of course does not prove that the transmutation is actually proceeding according to the exponential law; that is proved only for certain substances of which the half-periods amount to a few months, days or hours. Nevertheless we assume it, and multiply the reciprocal of λ_1 so measured by $\log_e 2$, and call the product the half-period. The values thus obtained are close to 1700 years for radium, agreeing well with the results of the method just above described; 4.7 billions of years for uranium I, agreeing with the result derived from the relative proportions of uranium and radium in the rocks; and 22 billions of years for thorium.

There are yet other ways of estimating half-periods, some of them very ingenious. Extremely short-lived substances require special methods. Thoron, a gas with the half-period of fifty-four seconds, is blown with a measured velocity through a tube along which various electrodes are placed for measuring its activity as it flows past them. Actinium A, of which the half-period is only .002 second, is projected upon the rim of a rapidly revolving wheel, and whirled past various instruments which measure its activity at successive points of its transit through space and time. The projection is due to a very simple but none the less striking natural phenomenon; when an alpha-particle is fired out of an atom of its parent-substance actinon, the residual particle—the atom of actinium A—rebounds or recoils like the

gun which fires a shell. The speed of this "recoil atom" is calculable, standing as it does to the speed of the ejected particle in the inverse ratio of their masses; and it has been utilized for measuring an excessively short half-period, that of RaC′, which amounts to only 10^{-6} second; a tube was oriented so that some of the recoil atoms flew along it, and their activity at various points of their flight was measured as in the case of thoron.[21]

Many of the half-periods, finally, are determined by analyzing the curves which represent the variation in time of the rays from continually-changing mixtures of growing and decaying substances: curves which presumably can be represented as sums of three, four or even more terms like the exponential terms in equation (6), not however independently known—that is to say, their coefficients and their exponents must be determined by inspecting the activity-curve itself and trying to build one like it. This operation sometimes requires a great deal of skill and discernment and intuition. It seems little short of marvelous that all the radioactive substances of the known series should have been recognized and their half-periods measured. That they have all been recognized there can be little doubt; for let us consider what it would imply if another substance lay undetected between (let us say) radon and radium A. There would have to be not one such substance but three, one of them emitting alpha-rays and the two others beta-rays—for otherwise the displacement-law of Fajans and Soddy would be broken. But if there were an undetected alpha-ray-emitting substance between radon and radium A, the atomic weight of the latter would be eight units below that of the former, instead of only four as we now suppose; and this difference of four units would follow step by step all the way down the radium series, ending in a to-be-expected value of 202 instead of 206; which would vitiate the excellent agreement between the latter figure and the observed atomic weight of the samples of element 82 contained in the uranium ores. The same argument can be used in the thorium series; in the actinium family the basis for the argument is lacking, but the parallelism between this and the other two families conduces to the same belief. It is all but certain, therefore, that the explorers of radioactivity have done their work so thoroughly that no substance yet remains unknown in the direct genealogical line from uranium I to radium G, nor in that from thorium to thorium D, nor between radioactinium and actinium D.

[21] This experiment was first performed by J. C. Jacobsen, and later by A. W. Barton, whose paper (*Phil. Mag.* (7) **2**, pp. 1275–1282; 1926) should be consulted for details. It is a very delicate one, especially as the atoms recoil because they have emitted not alpha-particles but electrons, which are comparatively light and are emitted with various speeds.

We turn to the rays themselves.

The alpha-rays are particles of mass $6.60 \cdot 10^{-25}$ gramme and positive charge $2e$ or $9.55 \cdot 10^{-10}$ electrostatic unit. The particles emitted from different radioactive substances differ, so far as we know, only in their initial speeds. The range of variation is astonishingly small; the slowest known alpha-particles issue from their sources (atoms of uranium I) with a speed of $1.423 \cdot 10^9$ cm/sec, the fastest[22] emerge from atoms of thorium C′ with a speed of $2.069 \cdot 10^9$ cm/sec. The differences in speed between different alpha-particles emerging from a substance are imperceptibly small.

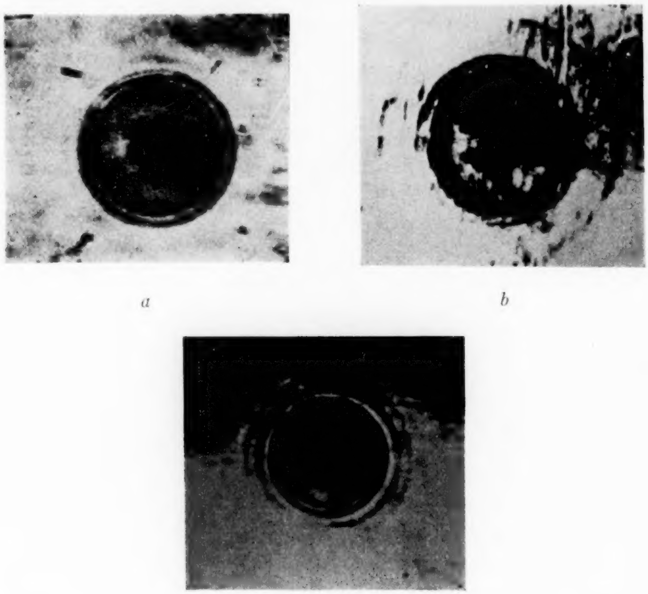
As a rule the speed of the alpha-rays from a substance is neither measured nor quoted directly; one measures by preference their *range in air*, a thing which can be defined because alpha-particles ionize air (and other substances) more readily the more slowly they are moving, until their speeds drop below 10^8 cm/sec and they suddenly cease to ionize altogether. Consequently, if alpha-rays shoot out from a bit of radioactive substance into environing matter, the concentration of the ions which they produce increases steadily and rapidly from the emitting substance outwards, up to a distance where it attains a sharp maximum and then suddenly falls to zero.[23] This distance is the range in the material in question; it is greater the faster the alpha-rays, varying as the cube of their initial speed. It is a property of the alpha-rays and not of the substance which emits them, and I should not have introduced it here but for a certain relation between ranges and half-periods, and as a pretext for showing some pictures of pleochroic haloes.

These haloes occur in certain ancient minerals, chiefly mica; they are systems of concentric spheres of discoloration, of which the pictures represent cross-sections. No one could imagine what they were when they were first discovered; but the explanation is simple and beautiful. Particles of uranium in some cases, of thorium in others, bubbles of radon in yet others, were caught ages ago and held in the points which were to become the centres of the haloes; the spheres of discoloration are the regions of maximum intensity of ionization, where the alpha-rays emitted from the central source were slowed down to their speed of optimum ionizing-power and were on the verge of

[22] Among the particles issuing from samples of thorium C and producing scintillations on fluorescent screens, very occasional ones (one in ten thousand, or fewer) have a much greater range than the rest, or than the characteristic particles of other substances. A few corpuscles of abnormally long range issue from samples of radium C. It is a controversial question whether these particles come from nuclei disintegrating in a rare and abnormal manner, or from nuclei struck and broken by alpha-particles ejected from other atoms, as sometimes happens. Even the published data are not all in accord, and it is unsafe to make further statements.

[23] *Introduction*, pp. 200–204.

ceasing from ionization altogether. The radius of the outer boundary of every such sphere is the range, in mica, of the kind of alpha-rays which caused it. All of the alpha-ray-emitting descendants of the initially-imprisoned substance form their individual spheres; in the cross-sections of the best haloes one can discern nearly all of the rings due to uranium I and its seven alpha-ray-emitting descendants on the direct line to radium G, or those of the seven members of the thorium



a *b* *c*

Fig. 3. Pleochroic haloes (B. Gudden, *ZS. f. Physik*)

a. Rings of UI and UII (innermost, merged into a single broad ring), Io, and Ra.
b. Rings of RaF (innermost), Rn, RaA and RaC′.
c. Rings of various substances of the uranium-radium series. Magnifications 665, 500, 480 respectively.

series which disintegrate in this way. There are no extra rings in these haloes, which strengthens the presumption that no radioactive substances in either series lie undetected. But there are also haloes of which the rings have not the proper radii to be identified with any known radiating substance. Are these possibly evidence for the prehistoric existence of others belonging to other series, all of which were too short-lived to survive into the days of scientific research, but disappeared with the dinosaur and the pterodactyl?

There is an interesting and important relation between the initial speeds of alpha-rays and the half-periods of the substances which emit them. One varies as an exceedingly high (negative) power of the other, so that when speed is plotted versus half-period upon logarithmic plotting-paper the resulting curve is a straight line; or, rather, three parallel straight lines, one for each of the three series. This remains true if we plot any power of the speed (for instance, the third) against

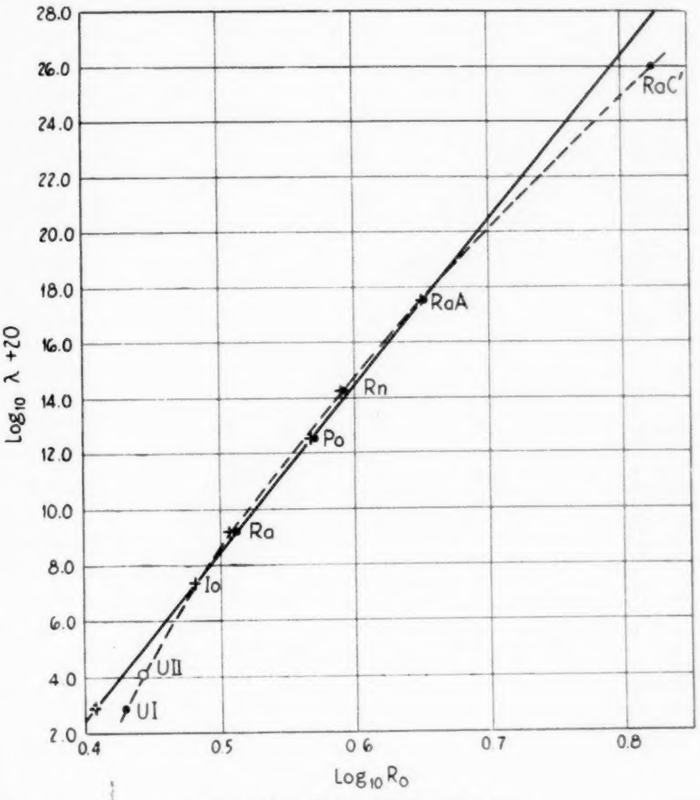


Fig. 4. The Geiger-Nuttall relation

Data for the uranium-radium series; the values for alpha-particle range denoted by dots are taken from pleochroic haloes, those marked by crosses from experimental data. The value of half-period for UII is not known, but is placed by interpolation upon the smooth curve. The straight line represents the best attainable approximation by a linear relation between logarithms of half-period and range; the smooth curve is that drawn by B. Gudden, from whom the data are taken.

any power of the half-period (for instance, the power -1), so that the logarithm of the range of the emitted particles varies linearly with the

logarithm of the disintegration-constant of the emitting substance. This is the way in which this, the "Geiger-Nuttall" relation, is usually expressed:

$$\log \lambda = A + B \log R. \tag{12}$$

The constant B is given (by Hevesy and Paneth) as 53.9 for all three series of radioactive substances, which signifies that the disintegration-constant varies as the fifty-fourth power of the range of the ejected particles! I do not know of any other relation between physical variables in which so high a power occurs; radioactivity, like astronomy, is the home of colossal numbers. The constant A varies from one series to another; it is given as -37.7 for the radium series.

The Geiger-Nuttall relation, like most simple formulae, is a mere approximation. For the radium series its degree of accuracy is illustrated by Fig. 4; the curve drawn through the various points is not quite a straight line. In the actinium series there is a jolt; the point for actinium X lies quite away from the place it should occupy on the straight line drawn to fit closest to the points for the other members, and in fact the half-period of AcX is shorter than that of RdAc, though its alpha-particles are slower. A straight line can be drawn to pass near the points for the remaining members of this series, and another to pass near the points for the descendants of thorium, about as successfully as the line in Fig. 4 fits the points for the radium family. Extending the line drawn for the thorium family to the value of the range for the fastest of all alpha-particles,[24] those of thorium C′, one obtains by extrapolation for the half-period of this substance the fantastically small value 10^{-11} second. There is no discernible prospect of verifying this by direct measurement, and in quoting it one should remember the risks of extrapolation.

An alpha-particle is a helium nucleus; when it acquires two electrons, the combination is a helium atom. Helium therefore is a daughter-substance of every radioactive substance which transmutes itself by emitting alpha-rays.

Passing over from alpha-rays to beta-rays, we take at once a great step backward from the clear to the obscure.

The great trouble arises from the fact that beta-rays are electrons, and electrons exist both in the atom-nuclei and in the electron-systems which surround them, or at least they come out of both localities. Whereas the emergence of an alpha-particle from a substance is a clear sign of the transmutation of one atom of that substance, the emergence of a beta-particle need not mean anything of the sort; it may

[24] Reservation being made for the particles mentioned in footnote 22.

simply mean that an alpha-particle, or a gamma-ray quantum, or a different beta-particle coming out of one atom-nucleus operated on its way out the expulsion of that electron from the outer electron-family of that atom or some other. To take one instance only: radium C and radioactinium both emit beta-rays and alpha-rays together, but in the former case there is as we have seen a dual transmutation, in the latter the beta-rays appear to be electrons torn out of the electron-shells surrounding the atom-nuclei as the alpha-particles pass by on their way out. Electrons have the same charge and the same mass, whatever their origin; although it is essential to distinguish how they originate in all these cases of beta-ray-emitting substances, there is no way to make the distinction except by performing experiments on distribution-in-speed of the beta-rays and invoking various theories, not always of the highest order of reliability, to interpret the results. This is the reason why, as Meitner says, the beta-rays actually emitted from self-transmuting nuclei "are the least clarified point in the entire problem of the radioactive transformations."

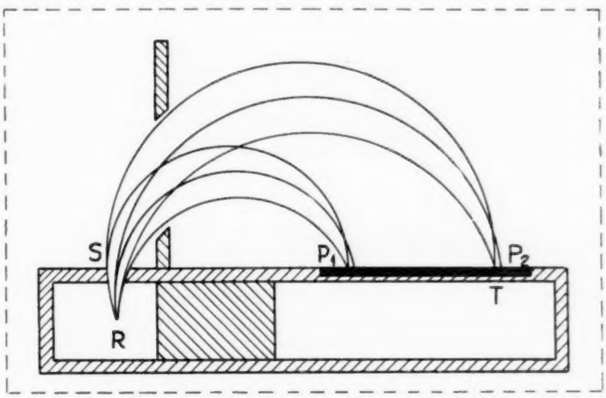


Fig. 5. Apparatus for photographing beta-ray spectra

(After C. D. Ellis and H. W. B. Skinner. Source at R, magnetic field and photographic plate perpendicular to plane of paper, which the plate intersects along P_1P_2.)

In attacking the beta-rays the first thing to do, and indeed the only thing which can be done by experiment, is to determine their distribution-in-speed—the function which gives the relative number of electrons issuing from the substance with speeds comprised between any preassigned limits. The process consists in isolating, by a proper system of narrow perforations and slits, a narrow beam or pencil of electrons, and applying to this pencil a magnetic field which bends the paths

of the electrons (see Fig. 5). The slower the electron, the more its trajectory is curved; if the beam comprises particles of more than a single speed, it is spread into a fan, and a photographic plate placed across the path of the fan records the "magnetic spectrum" of the electron-beam. If the beam comprises several groups of electrons, each with its own sharply marked and definite speed, each group falls upon a distinct part of the plate; if the slit limiting the beam is long and narrow, the groups form long and narrow discolored bands upon the plate, and these bands or "lines" constitute an electronic line-spectrum. The appearance of lines in a magnetic spectrum is taken as practically convincing evidence that the electrons in question issue from the circumnuclear electron-families of the atoms, not from their nuclei.

For this view there is direct evidence of a very convincing character: namely, that beta-ray spectra containing the same lines can be elicited from ordinary stable elements not undergoing transmutation, by the simple process of playing gamma-rays upon them from the radioactive substance in question. Take a sample of the substance, and envelop

Fig. 6. Part of the beta ray spectrum of radium B

(After Ellis and Skinner, *Proc. Roy. Soc.* Range 0.037 to 0.054 millions of equivalent volts.)

it in a metal sheath thick enough to stop all of the electrons or alpha-particles issuing from it. Some of the gamma-rays will pass through the sheath, for generally some of them (not necessarily all) are more penetrating than any other radiations which the substance can emit. Let these fall upon another piece of metal nearby; apply a magnetic field to the electrons expelled from this metal, or indeed to the electrons which the gamma-rays expel through the outer surface of the sheath enclosing their source; photograph the resulting spectrum. If the atomic number of the irradiated metal does not depart too far from that of the radioactive source—if for instance the irradiated metal is uranium or lead or platinum or tungsten—the spectrum of the electrons expelled from it will resemble the natural beta-ray spectrum of the source, closely enough so that strong lines of the one spectrum can obviously be identified with corresponding strong lines of the other. Corresponding lines in the two spectra may or may not coincide with

one another; that depends on the kind of metal irradiated; a prominent line in the spectrum of (for instance) radium B will be composed of electrons having energy somewhat greater than that of the electrons forming the corresponding line in the spectrum elicited from uranium, somewhat less than that for the corresponding line from platinum. But if the irradiated metal be isotopic with the substance into which the radioactive source is being transmuted, corresponding lines will be found to coincide with one another. One obtains a beta-ray spectrum having many lines in common with that of radium B, by allowing the gamma-rays to play upon and expel electrons from a piece of a metal isotopic with radium C—that is to say, bismuth.[25]

Whether one uses an atom-model or not, these facts suggest that some at least of the electrons emerging from a radioactive substance are hurled out by some sort of a secondary process operated upon the already-transmuted atoms by the accompanying gamma-rays, working in the same manner as they work upon atoms exposed to them outside. This suggestion becomes much more precise when the atom-model is invoked; for the contemporary model is designed to give a vivid explanation of the lines in the electronic spectra elicited by X-rays and gamma-rays playing upon the atoms of the stable metals.

Every such line is composed of electrons extracted from a particular group, in the circumnuclear electron-family of the atom, by radiation of a particular frequency. Think of the most tightly-bound electrons of all, the so-called *K*-electrons, to be imagined as lying or revolving closer than any of the others to the nucleus. Merely to extract one electron of this set, a definite amount of energy W_K must be imparted to the atom. Conceive a beam of radiation of frequency ν pouring over a multitude of similar atoms; to each it communicates either no energy at all, or else a definite amount of energy equal to $h\nu = 6.57 \cdot 10^{-27}\nu$. If this "quantum" unit of energy exceeds W_K, and if the radiation extracts a *K*-electron from an atom, the liberated electron will fly away with a kinetic energy equal to the excess of the imparted energy $h\nu$ over the extraction-energy or "binding-energy" W_K.

$$\text{(13)} \qquad \text{Kinetic Energy} = T = h\nu - W_K.$$

This equation determines the initial speed of the departing *K*-electrons.[26]

[25] *Introduction*, pp. 184–192; to this I refer also for reproductions of some very beautiful photographs of beta-ray spectra taken by J. Danysz and M. de Broglie.

[26] If the speed v of the electrons is inferior to $3 \cdot 10^9$ cm/sec, it is permissible to set for T the familiar expression $\frac{1}{2}mv^2$, putting for m the "rest-mass" $m_0 = 9 \cdot 10^{-28} g$ of the electron. Otherwise it is necessary to take account of the dependence of the mass of the electron upon its speed, preferably by using the formula derived from the

If there is but one frequency in the inflowing radiation, the spectrum of the emitted electrons will contain one line composed of what were formerly K-electrons. It will contain others, composed of electrons which originally belonged to other and less firmly-bound sets within the atoms. We distinguish, in order of decreasing binding-energy, K and L_{I} and L_{II} and L_{III} and M_{I} and M_{II} and M_{III} and M_{IV} and M_{V} and still further classes of electrons. The electron-spectrum due to radiation of a single frequency attacking atoms of a single kind comprises a line for each of these classes (apart from those, if any, for which the binding-energy exceeds the quantum-energy $h\nu$ so that the radiation cannot detach them) and the speed of the electrons composing each line is determined by an equation like (13), with the appropriate extraction-energy $W_{L\text{I}}$ or $W_{L\text{II}}$ or whichever it may be inserted in place of W_K. If there is more than one frequency in the incident radiation, each produces its own system of lines. These statements are proved, and the binding-energies are determined for all the classes of electrons and most of the kinds of metallic atoms, by irradiating metals with X-rays of which the frequencies are known, for they can be separately measured.[27] To ascertain the binding-energy of, let us say, the L_{II} electrons of platinum, one has only to look into the standard tables.

Now we have seen already that the physical and chemical properties of each radioactive substance, so far as they are known, are almost exactly like those of its stable isotope (if it has one); and with this rule the resemblance between the beta-ray spectrum of a radioactive substance and the electronic spectrum which its gamma-rays elicit from its stable isotope most admirably conforms. When a line in the former spectrum obviously corresponds to a line in the latter, both presumably are composed of electrons extracted from the same level by the same radiation. The same gamma-rays are working upon atoms isotopic with one another, and therefore endowed with electron-

Theory of Relativity, to wit:

$$T = m_0c^2\left[\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right] = h\nu - W_K$$

X-rays generated by artificial means never have frequencies so high that the electrons which they expel move rapidly enough for the simple substitution $T = \frac{1}{2}mv^2$ to be inadequate; but the frequencies of some of the gamma-rays are so great that the electrons which they extract even from the K-layers of massive atoms depart with speeds much exceeding $3 \cdot 10^9$ cm/sec. J. Thibaud has made direct measurements of a certain gamma-ray frequency and of the speed of the electrons which it ejects from a certain group of known extraction-energy, which are compatible with one another and with equation (13) if the relativity-formula for T is used, but decidedly incompatible if T be set equal to $\frac{1}{2}m_0v^2$ or to the once well-known expression derived by Abraham (J. Thibaud, *L'effet photoélectrique composé;* Paris (Masson) 1926).

[27] *Introduction*, pp. 192–195, 273–282.

families classified into identical classes with identical binding-energies. There is an evident difference; the atoms in the latter case are ionized by radiation poured upon them from without, in the latter by processes which occur within their own nuclei. (Whether in the latter case a wave-train does actually leave a nucleus, and enjoy a real existence during the brief time before it reaches the circumnuclear electron which it is destined to eject, is a question to which it is not easy to give a

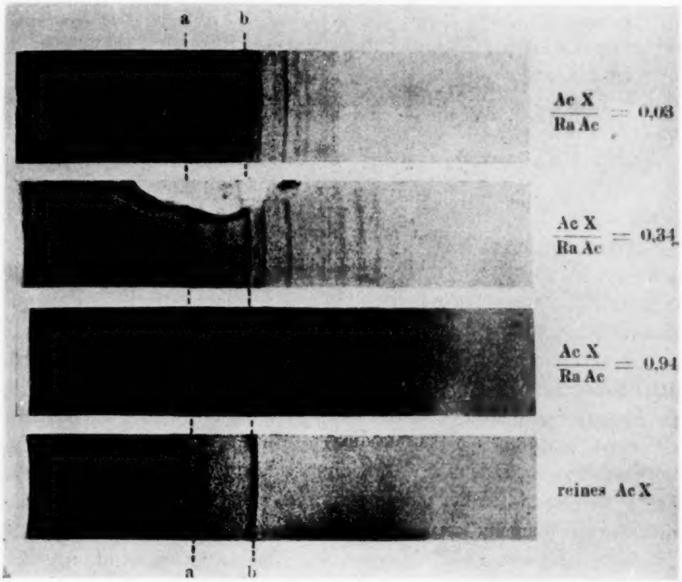


Fig. 7. Beta ray spectra of radioactinium and actinium X

(After O. Hahn and L. Meitner, *ZS. f. Physik.* The three upper pictures represent portions of the beta-ray spectrum photographed respectively a few hours, 6 days, and 20 days after the preparation of a pure sample of RdAc, in which the daughter-substance AcX was steadily growing; the lowest, the corresponding portion of the spectrum of a sample of AcX with its descendants. The lines which diminish in intensity from top to bottom are those of RdAc, those which increase belong to AcX and its descendants (note especially the lines marked *a* and *b*).)

sensible answer!) But the difference does not affect the energies of the ejected electrons; only their number, for, as seems natural enough, the beta-rays expelled from atoms of which the nuclei are emitting gamma-rays are much more abundant than those which an equal amount of gamma-radiation extracts from atoms on which it falls from without. Corresponding electron-groups have the same energy.

This is why we know in some cases, and suppose in the others, that when the electrons issuing from a radioactive substance constitute the lines of a line-spectrum they are not themselves coming from the nuclei; they are merely the signs of gamma-rays coming from the nuclei.

This discovery disposes of one potential objection to the displacement-law of Fajans and Soddy. Radioactinium (for instance) is a substance which emits alpha-rays and passes over into a substance two steps farther down the procession of the elements, as the displacement-law requires; but it also emits beta-rays, and since no alternative product one step farther up the procession has been discovered, the displacement-law would be gravely threatened if it were necessary to suppose that these come from the nuclei. There is no such necessity; and since the beta-rays display a line-spectrum, it is intrinsically all the more likely that they come from the circumnuclear electron-shells.

En revanche the character of the thus-far-analyzed beta-ray line-spectra makes it all the more difficult to understand what becomes of the electrons which must truly be emitted from the nuclei, in the transmutations in which the daughter-substance is displaced one step up the procession from its parent. When a substance is undergoing a transmutation of the other kind, the alpha-particles which its atoms emit all have very nearly the same speed. One would certainly expect that the electrons emitted from the nuclei of all atoms of radium B at their instants of transmutation emerge with the same speed. If so, they should compose a sharp line in the beta-ray spectrum of radium B. Now there are certainly some lines in this particularly rich spectrum which have not yet been definitely and exactly explained by the theory which I described before; but it appears that none of them is very prominent, and most of the experts refuse to admit that any one of them is composed of electrons coming forth direct and unretarded from the nucleus. There are other substances which display beta-ray spectra comprising but a few lines, one of which some physicists believe to contain the nuclear electrons.

If the nuclear electrons are not to be assigned to the lines, there remains but one alternative; they must be identified with the electrons composing the continuous beta-ray spectrum which underlies the lines and intervenes between them. The best way to study this spectrum is to dispense with the photographic plate, and set a Faraday-chamber to receive the electrons, with its aperture somewhere in the plane which the plate formerly occupied; if then the magnetic field is continuously varied, the spectrum slides across the aperture, and at each particular value of the fieldstrength the electrons of a particular limited speed-range pass into the chamber and are counted (more precisely, the total charge which they bear is measured, which comes to the same thing).

Curves obtained in this way are copied in Fig. 8. The peaks are the traces of lines (not so many as a photograph would show, for the method in this respect is not so delicate) rising up not from the zero-level but from a smooth sweeping curve, carried (hypothetically) in

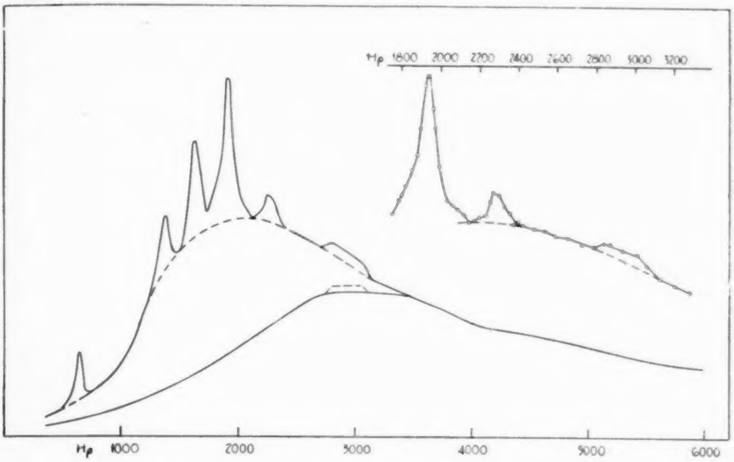


Fig. 8. Beta-ray spectra measured with a Faraday chamber
(Lower curve for RaC, upper curves for RaB + RaC. After R. W. Gurney.)

dashes across the base of the peaks. This is the distribution-curve of the electrons forming the continuous spectrum; integrating it, one obtains the total number of these electrons.[28] This number has been measured for radium B and radium C by Gurney; it amounts to somewhat more than one electron per self-transmuting atom.[29] A much smaller number, had such a one been found, would have rendered untenable the notion that all the nuclear electrons go into the continuous spectrum; the result proves that there is no such obstacle, not at least in these cases. The beta-ray spectrum of radium E consists of a single diffuse band; there are no lines. Emeléus counted the emitted electrons and found a value equivalent to 1.1 electrons per self-transmuting atom.[30] Perhaps then it is a quality of the nuclear

[28] This statement is not exact. The curves may be transformed into true distribution-curves, resembling them but not identically like them, by processes involving allowances for the geometry of the apparatus. The area under these true curves must then be found by integration, and gives the total charge borne by the electrons, the quotient of which by e is the desired number of electrons.

[29] R. W. Gurney, *Proc. Roy. Soc.*, **A 109**, pp. 540–561 (1925).

[30] K. G. Emeléus, *Proc. Camb. Phil. Soc.*, **22**, pp. 400–404 (1924). It is frequently pointed out that RaE emits no perceptible gamma-rays, a fact which makes it seem additionally probable that all the electrons which it emits come from the nuclei. This does not prove anything, as it is conceivable that gamma-rays are emitted which extract electrons from the electron-layers with such efficiency that no appreciable fraction of them escapes unconverted.

electrons that they issue from their atoms with widely and irregularly scattered speeds. If this is true, the presumption is that they escape from the nuclei with equal speeds, the differences resulting from experiences of theirs during the transit through the circumnuclear electron-shells. But it must not be forgotten that a continuous electronic spectrum appears together with the lines, when the gamma-rays from a radioactive substance fall upon one of its stable isotopes; and some allowance must certainly be made for this.

Refreshingly in contrast with the status of this perplexing question is the condition of another, for years the subject of a fervid controversy. Are the gamma-rays from a self-transmuting atom emitted before or after the transmutation occurs? There is only one way of settling this question, and perhaps the question itself ought to be so phrased as to bring this way into prominence. Granting the theory of beta-ray line-spectra expounded in these pages, and granting that certain lines in a certain spectrum have been recognized as being composed of electrons expelled by gamma-rays of one and the same frequency from various K, L, M classes in the circumnuclear electron-family, do the energy-values of these lines show that the electrons come from atoms as yet untransmuted, or from atoms which have already undergone their transmutation—from the atoms of the parent, or those of the daughter-substance? There is no forceful *a priori* reason for expecting either of these alternatives rather than the other; the question must be put to experiment.

If one knew with all desirable accuracy the frequency of the gamma-ray responsible for a particular set of lines, and the class of electrons contributing each line—if one knew for instance that a certain line is composed of K-electrons extracted by gamma-rays of a known frequency ν, one would measure the speed of these electrons, calculate their kinetic energy, subtract it from $h\nu$, identify the difference with the binding-energy W_K according to equation (11), and consult the standard tables to locate the element possessing that value of the extraction-energy for its K-electrons. But there are few gamma-rays of which the frequencies are independently known, and for these the values are not very accurate; so that this method is not generally available.

If however two lines are composed, the one of K-electrons and the other of L_{I} electrons ejected by gamma-rays of the same though unknown frequency, then the difference between the values of kinetic energy for the electrons of the two lines is equal to the difference between W_K and $W_{L\text{I}}$; and as this difference varies from element to element, one can consult the tables to locate the element for which the

difference between the K and the L_I extraction-energies agrees with the measured value. This is the usual method.

There is still another way, which may be explained by describing an experiment performed by C. D. Ellis and W. D. Wooster.[31] They enclosed a sample of radium B mixed with radium C in a rather thick-walled platinum tube, and deposited a thin layer of the same mixture upon the outer surface of the tube. The thin layer contributed the beta-ray spectrum of radium B and radium C. The beta-rays from the substances inside the tube were stopped by its walls, but the gamma-rays went through and expelled electrons from the platinum, which mingled with those from the covering film; so that upon the photographic plate there appeared side by side the spectrum-lines composed of electrons extracted from atoms of radium B and radium C by their own gamma-rays, and the spectrum-lines composed of electrons extracted from atoms of platinum by gamma-rays of the identical frequencies. Side by side there appeared, for instance, the lines due to K-electrons extracted by the same radiation from radium B and from platinum. The electrons from the radioactive substance had less energy than those from the platinum, for more had been spent in extracting them; the difference between the values of kinetic energy of the electrons was equal to the difference between the values of the K binding-energy for the atoms, with sign reversed; the K binding-energy for platinum is known, that of the other atom is calculated at once.[32]

The six or eight investigations, performed by these methods upon diverse substances by various physicists during the past two years, have all come to concordant results. The atoms from which the electrons of the beta-ray line-spectra are detached are the atoms of the daughter-substances; the gamma-rays are emitted, or at least they act (and it would be a daring person who would say that they exist for a while before they act!) after the transmutation occurs. The controversy

[31] *Proc. Camb. Phil. Soc.*, **23**, pp. 844–848 (1925). There are several important articles in this (November, 1925) number of the *Proceedings* which deal with the problem of the emission of gamma-rays, secondary X-rays, and electrons emitted from the nucleus or ejected from the circumnuclear family by these rays.

[32] All the methods require the observer to guess which lines are composed of electrons from the K-class, which of electrons from the L_I class, and so forth; and this is the major difficulty of the problem, for there is nothing intrinsically distinctive about the lines. In many cases, especially when there are several gamma-ray frequencies and a multitude of beta-ray lines, it is necessary to proceed by trial and error, assigning a line first to one class of electrons and then to another, and finally adopting the systematization which leaves the smallest number of lines unexplained or at odds. Sometimes only one out of the three L classes yields a perceptible number of electrons; there is a rule, which if general is very valuable, that when the product of h into the frequency of the gamma-ray exceeds the extraction-energies of all the L classes very greatly, then the L_I class is the only one out of which electrons enough are extracted to make a noticeable line.

is settled, and the triumphant side is that of which Meitner was the protagonist. Evidently we must conceive that the electron departing from a nucleus leaves it in a very unstable state, from which it speedily passes over into a comparatively though not absolutely stable state by one or a series of transitions, of which the gamma-rays are the manifestations.

We have still the gamma-rays to consider. Throughout this article I have taken it for granted that the gamma-rays are electromagnetic waves of definite frequencies. The evidence that they are electromagnetic waves has been known so long that it need not be rehearsed. The classical way of determining the frequency of such a wave is to measure its wave-length. With ordinary light-waves this is effected by dispersing them with a prism or diffracting them with a ruled grating. It used to be thought that these methods do not avail with X-rays, because of the shortness of their waves; natural crystal gratings, in which close-ordered files of atoms play the rôle of the rulings in artificial gratings, are used to diffract these. Applying the crystals to gamma-rays, one meets the same difficulty as the discoverers of X-ray met when they applied prisms and ruled gratings; the waves are mostly too short to be diffracted appreciably by natural crystals. The gamma-rays are spread out into a spectrum, and sometimes lines are discernible in the spectrum (Fig.9); but the line of shortest wave-length thus far measured (so far as I know) is at 0.052 Angstrom units or 52 X-units, and there are certainly many others at much shorter wave-lengths which the crystal spectroscope does not diffract far enough outward to be located. Recently people have renewed the attempt to measure wave-lengths of X-rays by the methods appropriate to visible light, and have attained values of astonishing accuracy; perhaps it is not too much to hope that a comparable advance in technique will bring the shortest gamma-rays into the scope of crystal gratings.

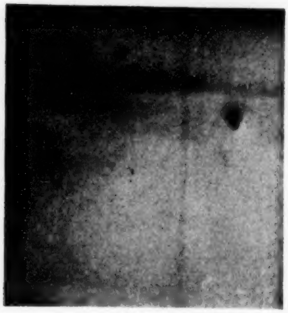


Fig. 9. Gamma-ray spectrum of radiothorium and thorium B

(Wavelengths of the lines 52X, 145X, 168X from right to left. After J. Thibaud, *l. c.* footnote 26.)

The usual method for estimating the frequencies of gamma-rays consists in guessing the class to which the electrons forming a beta-ray line originally belonged; taking its binding-energy from the tables; measuring the kinetic energy of the electrons forming the line; adding

it to the binding-energy, and dividing the quotient by h. This is in a sense the reverse of the usual process of ascertaining which is the element from which the beta-ray line-spectrum proceeds; and as a matter of fact the two have frequently been carried out as parts of one single investigation. The line-spectrum is photographed and its lines are measured, and then the student works over the data until he succeeds in setting up a hypothetical gamma-ray spectrum in which not too small a fraction of the beta-ray lines are explained by the action of not too great a number of gamma-ray frequencies upon the K and L and M classes of electrons, and reversely there is no too obtrusive case of a beta-ray line being predicted from his hypothetical gamma-spectrum and failing to appear.

For illustration I will quote some actual results. Meitner and Hahn located forty-nine lines in the beta-ray spectrum of radioactinium; among these, thirty-seven could be attributed to the action of one or another of twelve gamma-ray frequencies upon one or another of nine classes of electrons. In the spectrum of actinium X they observed twenty-nine lines and explained fourteen of them by postulating seven frequencies. Happily there are much more perspicuous cases. The spectrum of radium D consists of only a few lines—five, according to L. F. Curtiss,[33] whose measurements show that four may be supposed to consist of electrons ejected from the L_{I}, L_{II}, M_{I}, and N_{I} shells by a single gamma-radiation of wave-length 0.26A, while the energy of the electrons forming the fifth line is not perceptibly different from the quantum-energy $h\nu$ of the rays themselves. These last electrons may have been extracted from the outer layers of the atoms, where the binding-energy is so small that it makes but an insignificant deduction from the energy transferred to the electron. Another instance is that of radium itself, of which the three lines composing the beta-ray spectrum may be ascribed to a single gamma-ray of wave-length 0.066A expelling electrons from the K group, the L_{I} and the M_{I} group. Such cases as these are so simple that the theory in general and the wave-lengths calculated for the gamma-rays in particular are almost beyond all question.

Certain of the gamma-ray frequencies thus determined, and some which are directly measured with the crystal spectroscope, are found to agree with characteristic X-ray frequencies of the atoms whence they come. This is true of the solitary gamma-ray which is necessary and sufficient to explain the beta-ray spectrum of UX_1, and of two of the rays postulated by Meitner to account for the spectrum of RdAc

[33] L. F. Curtiss: *Phys. Rev.* (2), **27**, pp. 257–265 (1926). A previous investigation by L. Meitner (*ZS. f. Physik*, **11**, pp. 35–54; 1922) had led to substantially the same conclusion regarding the gamma-ray spectrum.

and two of those for AcX. This is precisely what was to be expected; for in the contemporary atom-model, the characteristic X-rays of an atom are conceived to arise from its circumnuclear electron-family, and to arise after and because an electron has been evicted from the family —a cause which the primary gamma-rays, or the alpha-rays or the electrons coming out of the nuclei, can themselves supply. The electrons expelled by these "secondary" gamma-rays or X-rays (the latter term is now preferred, in all cases where the identification can be surely made) are ejected as the *fourth* stage of a complicated process: first, the primary quantum or particle departs from the nucleus, then a tightly-bound electron is ejected from the electron-family, then a rearrangement of the remaining electrons brings about the emission of an X-ray, which in turn expels the loosely-bound electron. (It seems unlikely, as I intimated before, that the four stages are really separate; probably the passage from the initial state to the final takes place in a single operation, in a flash; but one does not see how to conceive that single operation without resolving it into four.) Since even the primary gamma-rays are emitted after the transmutation, the secondary X-rays *a fortiori* must come from the atoms of the daughter-substance; and this they do.[34]

The gamma-ray spectra thus far mapped out consist of from one to fourteen frequencies, not counting the secondary X-rays; the palm, in this respect, is awarded to radium C. The highest frequency thus far recorded is $5.4 \cdot 10^{20}$, corresponding to a wave-length of 5.57X (5.57×10^{-11} cm) and a quantum-energy amounting to $3.54 \cdot 10^{-6}$ erg or 2.22 millions of equivalent volts; it has twenty times the frequency of the highest X-ray known, and twenty times as great an energy in each quantum as is required to tear the most tightly-bound electron from the family of the most massive atom. It emerges from the nuclei of atoms which have just transmuted themselves out of radium C into radium C′. The fastest electrons forming a definitely-known line in a beta-ray line-spectrum occur in that of thorium C″; their speed amounts to 0.986 of that of light, their energy to almost $2.5 \cdot 10^6$ equivalent volts; but there are still faster ones in the continuous spectrum of radium C, which extends at least as far as to 0.998 of the speed of light. The energy of the alpha-particles of the various substances which emit them

[34] The strongest single piece of evidence is the measurement upon two radiations of radium B, performed with the crystal spectroscope by Rutherford and Wooster (*Proc. Camb. Phil. Soc.*, **23**, pp. 834–837; 1925) who found that the difference between the angles at which they were diffracted from the crystals agreed closely with that to be expected for two prominent X-ray lines of the *L* series of the daughter element (atomic number 83) and disagreed unmistakably with that to be expected for the parent element. This invalidated a contrary result obtained in 1914, which long had stood as an obstacle in the way of the conclusion that gamma-rays are emitted after the transmutation.

ranges from somewhat over four to somewhat under nine millions of equivalent volts. The greatest amount of energy which men have yet succeeded in loading upon a single charged particle or crowding into a single quantum of radiation lies well below the first million of equivalent volts; it still lay well below the first hundred thousand, ten years after radium was discovered. The step from the tens of thousands to the millions is a great one; this supplement voluntarily offered by Nature, transcending immensely the greatest amounts of energy which men can concentrate into a compact parcel, is chiefly responsible for the advances in the understanding of energy and matter which radioactivity made possible.

The advances have indeed been great. Consider what ensued from the discovery of the alpha-rays alone. With alpha-particles Rutherford explored the interiors of atoms, and the results of his explorations led him to the nuclear atom-model. The particles themselves he proved to be atom-nuclei of a certain element, and they established the amounts of electric charge which must be assigned to the atoms of that element and all the others. The nuclear atom-model in turn supplied Niels Bohr with the substructure of his theory; and Bohr's theory, together with the phenomena which it inspired men to seek and find, forms the half of contemporary physics. In the edifice of modern physical theory, the alpha-particle is the cornerstone. Had Nature not dispersed the radioactive substances through the rocks of the earth, had there not been one or two of them long-lived enough to survive and maintain a supply of their descendants until man arrived and became scientific—or if the faint outward signs of the radioactivity latent in the rocks had been overlooked, or having once been noticed had been left unstudied—in any of these cases, centuries more might have passed before a proper foundation was located for the edifice. That is the prime reason for honoring those who detected radioactivity, and then did not rest until they had brought it fully into the light. Theirs is an illustrious history, and one not without pathos; for some of those who had worked with the greatest zeal found themselves in later years the prey of a terrible and inexorable disease; like Prometheus in the myth, they were consumed for having brought benefits to the human race. Even yet the benefits which they gave have not been fully exploited; marvelous things may still be discovered, in the process of understanding the actions of the rays on living matter. But that will be another story, and a long one.

Dynamical Study of the Vowel Sounds
Part II

By IRVING B. CRANDALL

SYNOPSIS: Comparative studies based on oscillographic records of the principal characteristics of vowel, semi-vowel, and consonant sounds, have contributed much to an understanding of the mechanism of speech. Analyses of the frequency spectra of vowels show almost invariably two principal resonance peaks which fact is suggestive of a double resonator to produce them.

The present paper is concerned with the mechanism of the double resonator system and a mathematical treatment thereof. Based on the volume, shape and coupling of the resonating chambers, some models of cardboard, tube and plasticene were made, and with which some experimental tests in the production of vowels were carried out. The best success was had with the sound *ä* (father) while fair results were obtained with the sound *ō*, *ă* and *ĕ*.

INTRODUCTION

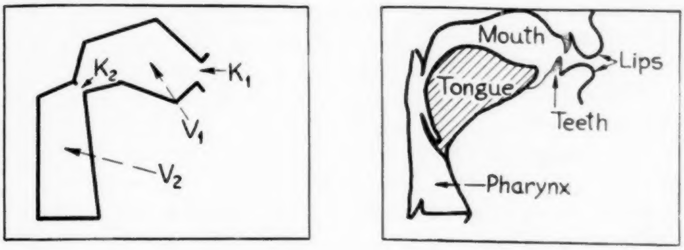
IN two earlier papers [1] a diagram has been given of the frequency spectra of the vowel sounds, based on analyses of a large number of accurate oscillographic records. In addition, there was given, in the second of these papers, a comparative study of the principal characteristics of vowel, semi-vowel and consonant sounds, and an account of certain studies made by other investigators whose methods and results have contributed to our understanding of the mechanism of speech.

Among the more original of recent contributions are those of Sir Richard Paget,[2] who has successfully employed multiple resonators to simulate almost all the vowel and consonant sounds. In getting together the material for the second paper from the Bell Laboratories, Paget's results for the *vowel sounds* were compared with ours only in a general way, and not in so detailed a manner as was followed in the discussion of consonant and semi-vowel sounds in that paper. It may be permissible to return to a consideration of the vowel sounds in the present paper, following Sir Richard Paget's idea of the double resonator as the instrument for vowel production. Indeed Sir Richard has pointed out to us that, since our own data on the spectra of the vowel sounds show almost invariably two principal resonance peaks, there must be a double resonator to produce them, thus harmonizing our results, at least for the male voices, with his own.

[1] I. B. Crandall and C. F. Sacia, "Dynamical Study of the Vowel Sounds," also "The Sounds of Speech," BELL SYSTEM TECHNICAL JOURNAL, III, 1924, p. 232; bid., IV, 1925, p. 586.

[2] *Proc. Roy. Soc.*, A102, 1923, p. 752; ibid., A106, 1924, p. 150.

Fig. 1*a* is a diagram of a double resonator. The volumes of the chambers are respectively V_1 and V_2; the conductivities [3] of the orifices are K_1 and K_2. In this structure the outer orifice corresponds to the mouth (see Fig. 1*b*), the outer cavity to the buccal cavity, the



1*a* 1*b*

Fig. 1*a* and 1*b*—Diagram of the mouth-pharynx system

inner orifice to the constriction between the soft palate and the back of the tongue, and the inner cavity to the pharynx. The source of sound in the back of the inner chamber is of course the glottis, or rather the periodic puffs of air to which the glottis gives rise, and we may remark that at resonance the apparatus is *driven at a node* (or pressure maximum) which is a condition for maximum efficiency. In Paget's models, a small opening was made at the back for the source of sound, which was a loosely stretched strip of rubber, mounted in a slit, and blown by an air stream. To be successful, in connection with the resonator model, in producing a vowel sound artificially, such a source must of course generate a sound whose fundamental is somewhere near that of the human larynx, and which has in addition a very extended range of harmonics; that is, for a bass voice, we should need a fundamental frequency of about 100, and additional sound energy scattered through the frequency range up to 4,000 or 5,000 cycles.[4]

[3] The average mass of air which surges to and fro in the orifice of a resonator is $\rho S^2/K$, in which ρ is the density of air, S the area of the orifice, and K the conductivity. K is a linear quantity, proportional to the width of the orifice, and is a measure of the ease of flow of fluid through it. It may be defined as the ratio of the (velocity) potential difference, between the two ends of the orifice, and the flux or current $(S\dot{\xi})$ flowing through the orifice.

[4] Sacia suggests that a source of sound giving a saw-toothed wave (rip saw tooth: one slanting and one vertical side) should be ideal for driving vowel resonators. (An experiment with such a device will be described later.) This wave shape corresponds to a fundamental and full retinue of harmonic tones, and should be of service in many ways in acoustic experiments.

Physical Features of the Mouth-Pharynx System

It is a curious fact that most of our data on the shape of the mouth cavities, position of the tongue, etc., for producing the different vowel sounds have been obtained by students of phonetics. There are of course excellent drawings of the mouth structure, in a few typical positions, given in the literature of anatomy; but for the finer differences, from one vowel sound to the next, we must rely on other sources. I know of no determination, for example, of the actual volumes of the mouth and pharynx, in any position for a typical individual, nor have I succeeded, by consulting anatomical experts, in obtaining the desired data.

In Fig. 2, there are shown certain conventional drawings, in median section, of the human mouth-pharynx region. These are taken from Rippmann's "Elements of Phonetics" (London, Dent, 1914) and were taken in turn by Rippmann from an article by Dr. R. J. Lloyd. In

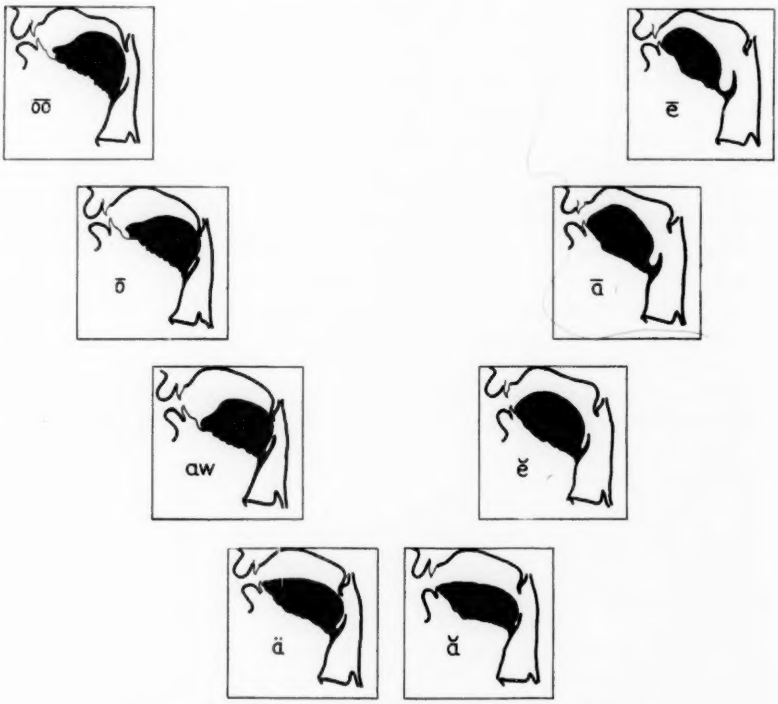


Fig. 2—Diagrams of vocal cavities for various vowel sounds

drawing conclusions from such diagrams as these, we must take care, of course, to use only the broadest features revealed by the series.[5]

It is evident that for the sounds on the left leg of the usual triangle (Fig. 3) (with the exception of short *u*), the inner orifice (that between the back of the tongue and the soft palate) is much constricted, and we

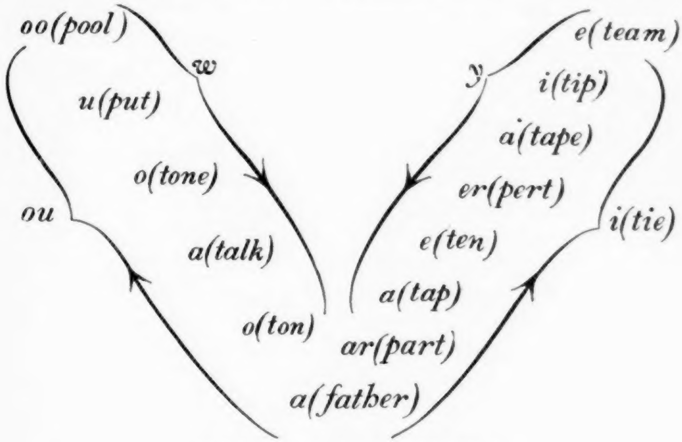


Fig. 3—Conventional vowel triangle

have here a loosely-coupled system to deal with. Also, due to the rearward position of most of the tongue structure, the mouth cavity appears larger than the pharynx. Here we must realize the horizontal width of these cavities, as well as their vertical extent. For the sounds on the right of the triangle, the tongue goes forward in such a way that the front cavity becomes the smaller of the two, and the connecting (inner) orifice becomes larger; the system then becomes closely coupled.

For some of the sounds it is not a difficult matter to get fair values for the conductivity of the mouth opening; this is approximately a circle, or an ellipse of moderate eccentricity in these cases. (The con-

[5] I understand that Prof. G. Oscar Russell, Director of the Phonetics Laboratory, Ohio State University, has made a remarkable series of clear X-ray photographs of tongue and mouth positions, for the various vowel sounds, some of which he has kindly shown me. He has worked out a special technique for making these pictures, and is now engaged in a thorough study of them, which will ultimately be published in monograph form. Unfortunately it is not possible to reproduce the pictures here; but it may be stated that the series follows (but in a more systematic way) the general course exhibited by the Rippmann diagrams shown in Fig. 2 of the present paper. The comparison between the results sketched in the present paper for the mouth cavities and results later to be published by Professor Russell should make a most interesting study.

ductivity of the circle is its diameter; we may take the conductivity of the ellipse as roughly equal to that of the circle of equal area.) In some cases, howevei (as for example, long $\bar{e}$), where teeth and lips are nearly closed together, the conductivity is certainly less than it appears on merely viewing the opening between the lips; hence a smaller value must be used. The conductivity of the inner orifice is even more uncertain, but in getting at this we are aided to some extent by a theoretical principle which will be given later. The diagrams at least offer some guidance in placing the various conductivities in thei order of relative magnitude.

The most serious lack of data relates to the volumes V_1 and V_2. I have made attempts to fill the mouth with water, and then measure this volumetrically, but of course this gives no hint of the volume of the pharynx. From these experiments, and other considerations, it seems that for an adult male the total volume $V_1 + V_2$ should be something over 100 cm.^{3}, and *nearly constant* for all the vowel sounds. That is to say, the change in V_1 and V_2 consists largely in a shift of volume from V_1 to V_2 (or vice versa) by the movement of the tongue; a proposition not so unreasonable anatomically, because competent advice states that a muscle, in taking its various shapes, preserves the same volume. Finally one would expect a somewhat larger total volume with the mouth wide open, for certain sounds, but this is partially compensated by the flattening of the cheeks in that position.

For the purposes of this study we shall consider $V_1 + V_2 \doteq 120$ cm.^{3} as one of the given data. But it may be stated in passing that these volumes should be much more accurately determined, preferably by anatomical experts.

It would be interesting to compare the results we shall obtain, for the dimensions of the resonator systems, with the actual data of Paget's resonators. But, on account of the four variables involved (K_1, K_2, V_1, V_2), there is no solution of a given case that is unique—that is to say, there are several combinations of different elements possible which will produce a given pair of natural frequencies. Hence such comparisons would often tell us little. Besides, in most cases it is impossible, from the figures given by Paget, to determine the sizes of his resonators, though their shapes are well shown in his drawings. Paget sometimes frankly imitated the structure of the mouth-pharynx system—not necessarily to scale—but sometimes, as in producing double (uncoupled) resonance by resonators in parallel, his models bore no relation to the structure of the natural system.

Spectra of the Vowel Sounds

We shall take as fundamental data the average spectra of the vowel sounds (for male voices) as given by the writer's previous work with C. F. Sacia, and as given in Sir Richard Paget's chart. We thus treat Sir Richard's data as if they had been obtained analytically, and not synthetically, for the sake of taking the mean values of the two most complete series of data available, to get a better basis for calculation.

The two principal resonant frequencies for each sound are given in Table I. The lower characteristic frequency is denoted by $\omega_1/2\pi$; the other by $\omega_2/2\pi$. These characteristic frequencies are also shown in the chart of Fig. 4.

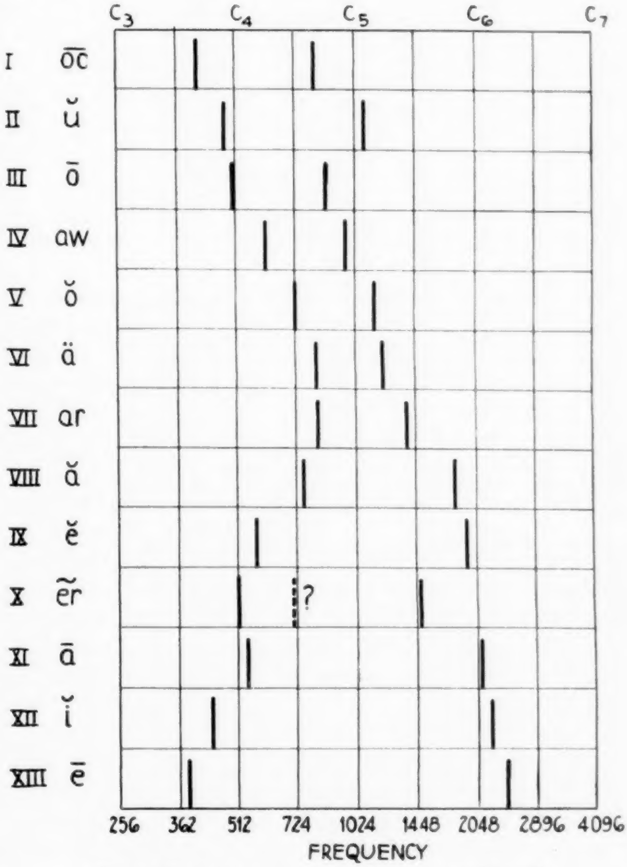


Fig. 4—Principal resonant frequencies, vowel sounds

TABLE I

NATURAL OR CHARACTERISTIC FREQUENCIES OF THE VOWEL SOUNDS

(*Male Voices*)

Sound	$\omega_1/2\pi$				$\omega_2/2\pi$			
	Crandall and Sacia	Paget (centered about)	Mean	Equiv. pitch	Crandall and Sacia	Paget (centered about)	Mean	Equiv. pitch
I. *oo* (p*oo*l) . . .	431	383	407	$G_3\#$	861	724	793	$G_4\#$
II. *u* (p*u*t) . . .	575	362	473	B_3-	1,149	966	1,058	C_5+
III. *o* (t*o*ne) . .	575	430	502	C_4-	912	790	851	A_4
IV. *a* (t*a*lk) . . .	645	558	602	$D_4\#$	1,024*	886	955	B_4
V. *o* (t*o*n) . . .	724	703†	713	$F_4\#-$	1,218	1,116†	1,167	D_5
VI. *a* (f*a*ther) .	861	790	825	$G_4\#+$	1,149	1,254	1,202	$D_5\#$
VII. *ar* (p*ar*t) . .	861	767	814	$G_4\#$	1,290	1,491	1,390	F_5+
VIII. *a* (t*a*p) . . .	813	703	758	G_4-	1,825	1,824	1,825	$A_5\#$
IX. *e* (t*e*n)	609	527	568	D_4-	1,825	1,932	1,879	B_5-
X. *er* (p*er*t) . . .	{ 542 700‡	470	{ 506 700	C_4	,448	1,534	1,491	G_5-
XI. *a* (t*a*pe) . .	609	470	540	$C_4\#$	2,048	2,169	2,108	C_6+
XII. *i* (t*i*p)	512	362	437	A_3	2,170	2,298	2,234	$C_6\#+$
XIII. *e* (t*e*am) . .	431	332	381	G_3	2,435	2,434	2,435	$D_6\#+$

* Poorly resolved, in our charts.

† In Paget's notation, for the sound *o* as in *not*.

‡ Considering *er* to have triple resonance.

The main resonances of most of these sounds are so pronounced that it is not at all difficult to take the correct data from the original charts, gnoring the less-essential minor peaks. In only one case (*a* as in *talk*) does our original chart fail to resolve the two principal peaks, but they are partially resolved even in this case, so that there is no great uncertainty in the figure given. In the case of the sound *er*, a third frequency is shown in the diagram. Reasons will be given later for considering this sound to be produced by a system of three degrees of freedom.

MECHANISM OF THE DOUBLE RESONATOR SYSTEM

There are two ways of studying the action of the double resonator at its resonant frequencies. If we drive the back of the inner chamber with a source of prescribed motion, then the greatest motion in the orifices will be obtained when the driving point impedance of the system *as viewed from the back of the inner chamber* is infinite. Or, equally, if we drive the system (by sound waves, say) from the front orifice, then the greatest motion will be obtained for those frequencies for which the driving point impedance of the system *as viewed from without* is zero. By either method we should be able to deduce the natural frequencies of the system; the second method is chosen here because it involves less labor.

In Rayleigh (II, p. 191, eq. 12) it is shown that the natural frequencies of a double resonator of the type described are the roots ω_1, ω_2, of

$$\omega^4 - \omega^2(n_1^2 + n_2^2 + n_{12}^2) + n_1^2 n_2^2 = 0, \tag{1}$$

in which

$n_1 = c\sqrt{\dfrac{K_1}{V_1}}$, the natural frequency of the outer resonator, with inner orifice closed;

$n_2 = c\sqrt{\dfrac{K_2}{V_2}}$, the natural frequency of the inner resonator alone;

$n_{12} = c\sqrt{\dfrac{K_2}{V_1}}$,

and c is the velocity o[illegible]d. Equation (1) is easily obtained by writing the equations of motion of the system, for zero applied forces and zero damping, and placing the determinant of the coefficients of the amplitudes or velocities equal to zero.[6] (This is equivalent to placing the driving point impedance, as viewed from the front orifice, equal to zero.) If $n_{12} \doteq 0$ (the case of a very constricted inner orifice), the roots of (1) are simply n_1, n_2.

We neglect damping in the system in order to get an easily-managed solution for the natural frequencies. Damping arises in two ways: (1) from sound absorption by the soft (tissue) lining of the cavities, and (2) by radiation from the mouth. Both are very variable, that due to radiation particularly so because of the considerable change in size of the mouth opening from one vowel sound to another. A great deal can be learned of the mechanism of the system by studying only the natural frequencies, and although it is not entirely impracticable to solve the problem with an allowance for radiation damping, we shall ignore this here.

The general procedure in this study will be to take as known from the vowel spectra the actual natural frequencies ω_1, ω_2 of the system, and to find the most reasonable values for the four quantities K_1, K_2, V_1, V_2, in order that these natural frequencies may result. We thus reconstruct the hypothetical resonator, or throat-mouth system which produces the vowel sounds. If we take

$$n_{12} = c\sqrt{\frac{K_2}{V_1}} = n_1\sqrt{\mu}, \qquad \left(\mu = \frac{K_2}{K_1}\right), \tag{2}$$

[6] A typical solution of a double resonator problem is given in the author's "Theory of Vibrating Systems and Sound," Van Nostrand (1926), pp. 59–64. The double resonator as a sound amplifier is discussed by E. T. Paris, *Science Progress*, XX, No. 77 (1925), p. 68.

we may rewrite (1) as

$$\omega^4 - \omega^2[n_1^2(1 + \mu) + n_2] + n_1^2 n_2^2 = 0. \tag{1a}$$

If it were not for μ, we could determine from (1a) the ratios K_1/V_1 and K_2/V_2 from the known data ω_1, ω_2. As will appear later, we can make reasonable assumptions with regard to μ; but it is obvious that even then two further assumptions are required to fix K_1, K_2, V_1, V_2 in absolute value. These we supply by assuming a fixed total volume $V_1 + V_2$ for the system, and a certain conductivity K_1 for the mouth opening, which is the most easily observed element of the system.

Proceeding in the manner outlined, it will be possible to take the series of the vowel sounds and fit to each sound a doubly resonant system such that the whole series forms a more or less coherent group.

The following is an outline of the type of calculations required. If we write, from (1a),

$$\begin{aligned} n_1^2(1 + \mu) + n_2^2 &= \omega_1^2 + \omega_2^2, \\ n_1^2 n_2^2 &= \omega_1^2 \omega_2^2, \end{aligned} \tag{3}$$

and eliminate n_2^2, we have

$$\left.\begin{matrix} n_1^2 \\ n_1'^2 \end{matrix}\right\} = \frac{\omega_1^2 + \omega_2^2 \pm \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2(1 + \mu)}; \tag{4}$$

also, if we eliminate n_1^2, we have

$$\left.\begin{matrix} n_2^2 \\ n_2'^2 \end{matrix}\right\} = \frac{\omega_1^2 + \omega_2^2 \mp \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2}. \tag{4a}$$

In these equations it will be noted that (n_1^2, n_2^2) $(n_1'^2, n_2'^2)$ each represent possible combinations of simple resonators which will give, on coupling, the observed frequencies ω_1, ω_2. In other words, for given (comparable) conductivities K_1, K_2, of the two orifices, the outer resonator may be small, and the inner resonator large ($V_1 < V_2$), corresponding to the (separate) natural frequencies

$$\begin{aligned} n_1^2 &\equiv c^2\frac{K_1}{V_1} = \frac{\omega_1^2 + \omega_2^2 + \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2(1 + \mu)}, \\ n_2^2 &\equiv c^2\frac{K_2}{V_2} = \frac{\omega_1^2 + \omega_2^2 - \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2}, \\ n_1^2 &> n_2^2; \end{aligned} \tag{5}$$

or, if $V_1 > V_2$, we must apply the other pair of equations

$$n_1'^2 \equiv c^2\frac{K_1}{V_1} < n_2'^2 \equiv c^2\frac{K_2}{V_2} \tag{5a}$$

using the lower signs in (4) and (4*a*). Thus in reconstructing the resonator cavities from the vowel data, we must take care to use, for each particular vowel, that pair of solutions (n_1, n_2, or n_1', n_2') which places the front and rear cavities in correct order for relative size. From the discussion given above of the data on position of the tongue, sections of the cavities, etc., the application of this principle is a relatively easy matter.

The matter of fixing the coupling factor is not so straightforward. For the loosely coupled systems (*oo* to *ar*, the vowels on the left leg of the triangle, Fig. 3), it appears that the maximum allowable coupling factors μ (that is, the values of μ for which the radicals in (4) and (4*a*) vanish) are so small that it seems reasonable to adopt them forthwith.[7] In these cases we have the single solution

$$\left.\begin{aligned} n_1^2 &= \frac{\omega_1^2 + \omega_2^2}{2(1+\mu)}, \\ n_2^2 &= \frac{\omega_1^2 + \omega_2^2}{2}, \qquad n_1^2 < n_2^2; \\ \mu &= \frac{(\omega_1^2 - \omega_2^2)^2}{4\omega_1^2\omega_2^2} = \frac{K_2}{K_1}. \end{aligned}\right\} \tag{6}$$

In this situation (since the ratio V_2/V_1 is fixed if n_1^2/n_2^2 and K_2/K_1 are fixed) all the quantities V_1, V_2, K_1, K_2 are determinate as soon as we fix either K_1 or $V_1 + V_2$. The practice followed will be to set a value for K_1 and check this by noting the value of $V_1 + V_2$ to which it leads; thus by trial and error the most reasonable values for the resonator constants for the loosely coupled systems can be found. Incidentally, we shall note in all these cases that the solution requires V_1 to be larger than V_2.

The vowel short *ă* marks the transition between the loosely-coupled systems already considered and the closely-coupled systems for the sounds from short *ĕ* to long *ē* on the right leg of the triangle. Short *ĕ* is also the first vowel sound of the series to have a high frequency resonance of frequency greater than 1,500 cycles. We might be in a

[7] These values of the coupling factors are not inconsistent with the diagrams of the mouth cavities shown in Fig. 2. Aside from complicating the calculations, the effect of taking still smaller values for μ (keeping K_1 constant) is merely to lower V_2 in proportion as K_2 is decreased. For example, taking $\mu = \mu$ max. for the sound *aw*, we arrive at the solution $V_1 = 119$ cu. cm., $V_2 = 22$ cu. cm., if $K_1 = 2.1$ cm. as given in Fig. 5. Now if we take $\mu = \frac{1}{2}\mu$ max., we get $V_1 = 121$, $V_2 = 10$ cu. cm. Thus no great change has been made in the total volume $V_1 + V_2$, except that we get a value for V_2 which seems unreasonably small. The most satisfactory course, in the case of the loosely coupled systems, is to use the maximum allowable coupling factors.

dilemma here, as to which pair of solutions (5 or 5*a*) to apply, since solutions are possible in which the two cavities V_1 and V_2 are of comparable size in this case. It is nearly certain, however, that the front cavity, V_1, is greater than V_2 in this case, but it is not certain that the highest possible value of μ ($\mu = 1$) is the one to use. A compromise was made, setting $\mu = .80$, and using equations (5*a*) for the solution. We shall see later that a resonator built according to these specifications performs sufficiently well to justify these assumptions. With this sound we have finished with equations (6) and (5*a*) and for the last time we have $V_1 > V_2$.

For the last 5 sounds (short *ĕ* to long *ē*) the maximum possible coupling factors range from 1.75 to 9.4; it has been found advisable to shade these and use factors ranging from 1.25 to 5.0. A choice now has to be made between solutions (5) and (5*a*); and since the tongue comes so far forward in these cases, we adopt at once the first solution, according to (5), which leads to the relation $V_1 < V_2$ in all these cases.

Discussion of the Results

The calculated results are shown in the chart, Fig. 5. Because of the speculative character of some of the assumptions made it is reasonable to call attention only to certain outstanding features of the chart. Among the first seven (loosely-coupled) systems the sound *u* (as in *put*), if placed *second*, would seem definitely out of order, because of the magnitude of the coupling factor, or (what is the same thing) the greater separation of the characteristic frequencies. There is no escape from the larger inner orifice for this system, and the effect which it produces. This sound simply does not conform to the habits of its (assumed) neighbors; otherwise the first seven sounds form a coherent group. In classifying short *u* Paget takes the dilemma by the horns, and places it *first*, that is, preceding all the other sounds of this group. This arrangement is adopted in Fig. 5.

There will be noticed in the chart a tendency to expand the total volume, $V_1 + V_2$, for the rounder and more open sounds. This is in a deliberate attempt to allow for the effect of opening the mouth a little wider in these cases.

The last 5 sounds (from short *ĕ* to long *ē*) form a fairly coherent group, except for the non-conforming member *er*. Paget places *er* preceding short *ă* in the series; it seems to the writer a hybrid of the short *ĕ* (or long *ā*) and the *r* sound, but its low frequency resonance (ca. 500) requires a large volume for either V_1 or V_2, and this can only be back of the tongue (V_2) because of the contraction of V_1 when the tip of the tongue is raised for the *r* sound. If we let $K_1 = 1.5$ cm., and

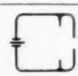
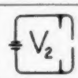
Sound	μ max.	μ used K_2/K_1	K_2	K_1	V_2+V_1	System (Schematic)	V_2	V_1
II ŭ (put)	.80	.80	.96	1.20	137		42	95
I o͞o (pool)	.50	.50	.45	.90	134		34	100
III ō (tone)	.31	.31	.45	1.50	146		28	118
IV a̤ (talk)	.23	.23	.48	2.10	141		22	119
V ŏ (ton)	.26	.26	.73	2.8	134		23	111
VI ä (father)	.15	.15	.52	3.5	126		15	111
VII ar (part)	32	.32	1.12	3.5	127		23	104
VIII ă (tap)	1.00	.80	2.0	2.5	123		23	100
X er (pert)	1.75	1.00	1.5	1.5	118		73	45
IX ĕ (ten)	2.27	1.25	2.25	1.8	117		77	40
XI ā (tape)	3.34	1.8	2.34	1.30	101		73	28
XII i (tip)	6.10	3.0	2.4	.80	102		80	22
XIII ē (team)	9.4	5.0	3.0	.60	104		83	21

Fig. 5—Schematic diagrams of doubly-resonant systems for vowel sounds

assume maximum coupling, i.e., $\mu = 1.75$, we get $V_1 = 98$ cu. cm. and $V_2 = 62$ cu. cm., which seems absurd; if we assumed for *er* a system of only *two* degrees of freedom, the most reasonable course would be to give μ a smaller value (say, unity) and solve on the basis that $V_2 > V_1$ which would give (if $K_1 = 1.5$) $V_1 = 45$ cu. cm., $V_2 = 73$ cu. cm., and $K_2 = 1.5$ cm. These data are entered (very tentatively) in Fig. 5; here again we revise the previous order, and place *er* between short *ă* and short *ĕ*.

It is not at all certain, however, from the spectra of the *er* sound (see chart, Fig. 13, in the paper "The Sounds of Speech") that it is produced by a system of only two degrees of freedom; the analyses of the female voices gave 3 definite peaks, and we note that when the tip of the tongue is raised, for this sound, there is a third cavity betewen the tongue and the lips which is doubtless significant. There will be noted, with a question mark, a third line (of frequency about 700, for the male voices) in the spectrum of *er* shown in Fig. 4. I have attempted, from the three lines shown in Fig. 4, and some simple assumptions regarding the volumes and conductivities, to obtain a rough solution, using 3 degrees of freedom for this sound; but none of these results are entered in the chart, because they appear to be unreasonable.[8]

No attempt has been made to subject the semi-vowel sounds (*l*, *ng*, *n*, *m*) to dynamical calculations. It is evident from their spectra (cf. "The Sounds of Speech") that they are produced by systems of three or four degrees of freedom, which is to be expected, if, in addition to mouth and pharynx, the tongue, naso-pharynx, or

[8] By trial and error it was hoped that some triply-resonant system could be found which would give the spectrum of *er*, as shown in Fig. 4. After solving more than a dozen of these systems, the best fit was one in which $V_1 = 31$, $V_2 = 63$, $V_3 = 31$ cu. cm.; $K_1 = K_2 = 1$ cm., $K_3 = \frac{1}{2}$ cm. The calculated frequencies for this system are 445, 890, and 1,520 cycles. The trouble with this solution is that the middle cavity (V_2, between the tongue and the roof of the mouth in this case) is the largest of the three, which does not seem reasonable. A model made to these specifications, and tried by the method described later, gave a sound something like *er* but not so satisfactorily that one could accept this as a solution. Consequently it is not entered in Fig. 5.

At first, in a number of these attempted solutions, the innermost chamber, V_3, was taken as the largest of the three. These all led to too great a separation of the two lower resonant frequencies to be acceptable.

The sound *er*, in addition to the three resonances about as shown in the chart, may contain a component of higher frequency; or it may be due to a progressive variation or modulation of the two principal frequencies shown in the chart. Some of Paget's results suggest this; and if this is so, it would be a most difficult vowel to imitate with a *fixed* resonator. It is possible that X-ray pictures may reveal some point hitherto overlooked in the mouth adjustment for this sound.

nasal cavities are brought into play. The calculations required would be too cumbersome for the present paper. It is rather a tribute to Paget's experimental skill that he was able to synthesize these more complicated sounds with resonators of more than two degrees of freedom and so arrive at their characteristics.

It is not thought that the calculations given herein suffer appreciably due to the omission of damping factors from the dynamical equations. It would be almost impossible to take correct values of damping constants from the speech spectra; there is a better chance of doing this from the records of the sounds themselves, but even so, they cannot be determined with anything like the precision of the natural frequencies.

To summarize the results, we have an idealized system of two degrees of freedom, loosely coupled for one group of the sounds, closely coupled for the remaining sounds, with fair indication of the transition between the two groups. We have the assumption of virtually constant total volume of the two cavities, and an indication of how this volume should be apportioned between them in most cases. We also have a rough determination in most cases of the conductivity of the inner orifice between the two cavities.

Some Experimental Tests

It would be of interest if we could now make models of all the systems considered, excite them in some suitable way, and establish their essential validity from the character of the sounds produced. This might seem unnecessary, on account of Sir Richard Paget's extended work; it seemed worth while, however, to attempt a few models, using cardboard tubes and plasticene for the structure.

The most success was had with the sound *a* (*fa*ther). A model was made to scale (Fig. 6), using the data of the chart—but of course

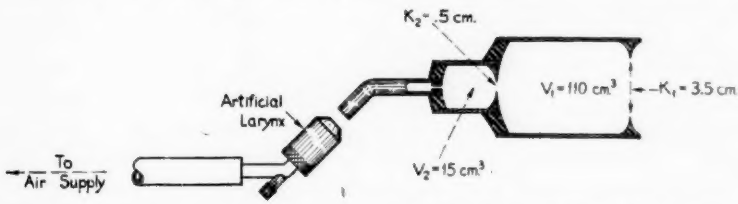


Fig. 6—Double resonator model for *ä*, and method of attaching artificial larynx

we should expect similar results from somewhat larger or smaller models, provided the ratios $K_1 : K_2 : V_1 : V_2$ were maintained; the chief point here is the variation in damping with the sizes of the orifices, and the requirement that any orifice should be smaller than

the mean dimension of the adjacent volume, in order that the usual resonator theory may apply.

The model when gently blown with a slow current of air through the small hole in the back gave a good *whispered ä*; but some difficulty was experienced in exciting it correctly for a *voiced ä*. It was first connected, at the rear, to an artificial larynx,[9] keeping the connecting hole small in order to preserve the dynamical characteristics of the main system. When the artificial larynx was blown (though it did not function well with the output orifice so small), a recognizable *voiced ä* was produced by the apparatus; but this was not as good as the whispered sound first described. (We have here the point made at the beginning: that the driving system, to imitate the vocal cords successfully, must give a low pitched tone, very rich in partials.) The artificial larynx was then replaced by a telephone receiver excited by the (rip) saw-toothed A.C. wave of 100 fundamental frequency, arranged by Mr. Sacia. A rather poor sustained *ä* sound resulted, quite deficient in volume, because of weak driving through the small hole in the back. Altogether, the artificial larynx, with its intermittent or variable excitation, came the nearest to producing a voiced *ä*; and the sound was similar to that produced by a person actually using the artificial larynx inserted in the side of the mouth, in the usual manner, for this sound.

Very fair results were also obtained with a model, built according to specifications, for the sound long *ō*. Models were next attempted for short *ă* and short *ĕ*. First, a model was made with two volumes $V_1 = 80$ cu. cm., $V_2 = 45$ cu. cm., and having the *three* openings K_1, K_2, and the hole in the rear of V_2 (for a cork fitting connecting the larynx) each about 2.5 cm. in diameter. It was thought that, when blown from the rear of V_2, it would give a recognizable short *ă* sound; and that when reversed, i.e., when the cork fitting was inserted in K_1 so that V_1 and V_2 became interchanged, it would give short *ĕ*. The result was that the sounds produced were nearly alike, and quite unsatisfactory in both cases! However, when the conductivities were modified, so that $K_1 = 2.5$, $K_2 = 2.0$, for short *ă*, and $K_1 = 2.0$, $K_2 = 2.5$ for short *ĕ*, the volumes being interchanged as before, the results were much better. As described here, the model for short *ĕ* approximates in dimensions the data entered in Fig. 5, but the model for short *ă* ($V_1 = 80$, $V_2 = 45$ cu. cm.) does not quite have the theoretical division of total volume (namely, $V_1 = 100$, $V_2 = 23$ cu. cm.) entered in the chart. The partition was therefore moved back, until this condition was obtained, with the result that the short *ă* sound was given at least as well as before.

[9] Previously described by H. Fletcher and C. E. Lane.

Attempts were also made at models for long *ā* and short *ĭ*, using the theoretical data. These seemed to give *whispered* sounds which suggested the true ones, but were not very satisfactory when excited by the artificial larynx. It is evidently more difficult to imitate the mouth structure by such simple means, when the outer conductivity (K_1, the orifice between lips and teeth) is small, and the inner orifice K_2 is large. And in addition it is likely that the artificial larynx does not supply sufficient high frequency energy to excite these sounds properly. There is also, of course, the difficulty of applying the simple resonator theory, when the conductivity of an orifice is comparable to one of the dimensions of the adjacent volume.

Conclusion

In this paper we have attempted to visualize the mechanism of the vowel sounds, on the basis of previous work, certain simple calculations, and a few rough experiments. It appears that the vowel sounds are usually produced by a double resonator system whose behavior in itself is thoroughly understood; but this does not by any means close the subject. A most interesting field of study remains in the excitation of the resonator system, to say nothing of the various factors which produce damping in the system itself.

We know from laboratory experiments that a reed (or a simple "squawker" made of rubber strip) is by itself a very poor imitation of the vocal cord apparatus. The artificial larynx, for example, will not vibrate properly unless a tube some 15 inches long is interposed between the "larynx" and the pressure reservoir by which it is blown. Correspondingly, we should expect the wind-pipe leading from the lungs to the human larynx to have a very important rôle in fixing the lower frequencies produced by the vocal cord apparatus. The mechanical problem indicated for study in this connection is the excitation of a reed-pipe with the reed at the distant end of the pipe, an inversion of the arrangement of ordinary wind instruments.

Consider the question of damping. In the apparatus used by J. Q. Stewart [10] (tuned electrical circuits excited by an interrupter) the damping could be systematically adjusted; this is the only case I know of, in experimenting with speech sounds, in which this adjustment was possible. In ordinary mechanical apparatus damping is difficult to control. Yet, damping is a significant element in the character of the constituent vibrations of either sustained or transient vowel sounds. For example, I have already pointed out [11] the close

[10] *Nature*, Sept. 2, 1922. "An Electrical Analogue of the Vocal Organs."

[11] "The Sounds of Speech," end of § V. Refer also to Records and Fig. 14 of that paper.

similarity between the spectra of *l* and long *ē*. In the semi-vowel *l* the characteristic high frequency (if viewed as a transient) decays much more rapidly than the corresponding vibration in the *ē* sound; this fact we have from the records themselves, but not from the frequency spectra. It may be that such phenomena as these will require a more definite adherence to the "transient" point of view in dealing with the vowel sounds, a matter previously discussed at some length.

The transitory or unstable qualities in the actual speech sounds almost defy imitation by mechanical means. There is, for example, the variation in fundamental frequency during the course of a vowel or semi-vowel sound which was pointed out in the paper "The Sounds of Speech." There is also the lengthening of the fundamental period for semi-vowels and voiced consonants as compared with vowel sounds; also the shortening of the fundamental cycle at the beginning of a voiced consonant.

Finally there is the question of classification of the speech sounds. We have already noted difficulties for some of the vowel sounds. It is likely that the vowel triangle or the arrangement of the vowels in a linear series will require modification. A satisfactory classification for all the sounds, from the dynamical standpoint, is at present an unsolved problem; but in conclusion one suggestion may be permissible. We might limit the application of the term "vowel sound" to those sounds which *can* be satisfactorily produced by the simple double resonator system. The more complicated vowel-like sounds (*l*, *ng*, *n*, *m* and possibly *r*) and some of the consonants can undoubtedly be related to systems of three or more degrees of freedom. A study of these systems is beyond the aims of the present paper; but it is to be hoped that such a study can be carried out, for the sake of the aid that mechanical theory offers in helping to visualize the mechanism of speech.

Radio Broadcast Coverage of City Areas [1]

By LLOYD ESPENSCHIED

SYNOPSIS: 1. Radio broadcasting involves a system of electrical distribution in which dependent relations exist between the transmitting station, the transmitting medium and the receiving station.

2. The attenuation and fading which attend the spreading out of broadcast waves are considered. The attenuation of overland transmission is shown to be, on the whole, very high and to vary over a wide range depending upon the terrain which is traversed. The distance at which the fading of signals occurs is found to be that at which the normal directly transmitted waves have become greatly attenuated and to depend upon the terrain traversed.

3. A field strength contour map is given of the measured distribution of waves broadcast by Station WEAF over the New York metropolitan area. A rough correlation is given between measured field strengths and the serviceability of the reception in yielding high grade reproduction. The range of a station as estimated in terms of year-round reliability is found to be relatively small. It becomes clear that the present radio broadcasting art is upon too low a power level and that higher powered stations are required if reliable year-round reception is to be had at distances as short even as 30 to 50 miles from the transmitting station.

4. The question of the preferred location of a transmitting station with respect to a city area is considered. It is shown that an antenna located upon a tall building may radiate poorly at certain wave-lengths and well at others. Surveys are presented of the distribution effected by an experimental transmitting station located in each of several suburban points. The locations are compared upon the basis of the "coverage" of receiving sets which they effect.

5. Finally, there is considered the relation which exists in respect to interference between a plurality of broadcast transmitting stations operating in the same service area. The importance of high selectivity in receiving sets is emphasized and there is given the measured selectivity characteristics for samples of a number of receiving sets.

IT is well recognized that the elements which comprise an electrical transmission system are required to function not simply as individual pieces of apparatus, but as integral parts of a whole. In the case of radio broadcasting, the absence of a common control of the two ends makes this over-all "systems" aspect less apparent than it is for wire systems.[2] Nevertheless a definite systems correlation is required between the broadcast transmitting station and each of the receivers served, as will be evident from the following:

1. The transmitter should put into the transmitting medium, without distortion and with the power called for by that medium, all of the wave-band components required and no others.

2. The transmitting medium should be capable of delivering to the receiver an undistorted wave band, reliably and stably, and with

[1] Presented at the New York Regional Meeting of the A. I. E. E., New York, N. Y., Nov. 11–12, 1926.

[2] Some other examples of such a "systems" relationship are given in "Application to Radio of Wire Transmission Engineering," published in the *Proceedings* of the Institute of Radio Engineers, October, 1922.

sufficient strength to enable the received waves to stand well above the level of the ever present interfering waves.

3. Finally, the receiving set should pass with the necessary volume all of the wave components required to reproduce the program signal and should sharply exclude all others.

The rapid apparatus development borne in by the vacuum tube has brought the art to the point where it is now physically possible to meet quite fully the terminal requirements. The apparatus development, in fact, has outstripped our knowledge of the transmitting medium itself, and we are now in the position of possessing apparatus possibilities without knowing very definitely the limitations and requirements placed upon their use by the intermediate link. Only within the last few years have methods become available for measuring radio transmission and thereby placing it upon a quantitative basis.

Such measuring means have been applied to the study of radio broadcast transmission from certain stations in New York City and in Washington, D. C. The earlier results of this measurement work have already been published. It is the purpose of the present paper to present results of a systematic study which has been made of the coverage which can be effected of the radio broadcast listeners of the New York metropolitan area and in so doing to portray something of the general systems requirements of radio broadcasting.

The Character of Radio Broadcast Transmission

The ideal law for broadcast distribution would be one whereby the transmitted waves are propagated at constant strength over the zone to be served and then fall abruptly to zero at the outer boundary. All receivers within the area would be treated to signals of equal strength and no interference would be caused in territories beyond.

The kind of law which nature has actually given us involves a rapid decadence in the strength of the waves as they are propagated over the service area, and then, instead of a sharp cut-off, a persistence to great distances at field strengths which, although often too low to be generally useful, are sufficient to cause interference in other service areas.

This situation is illustrated in Fig. 1. The upper curve shows the relation between intensity and distance; the lower portion, the interpretation of this curve in terms of areas of reception. The attenuation traced by the heavy line of the curve is that of the component of the radiation which is propagated directly along the earth's surface. It is this radiation which is ordinarily utilized for reliable broadcast reception. The shaded portions near the outer ends of this curve are

intended to indicate the appearance of variations in the signal intensity which occur at the greater distances, particularly at night, and which are known as "fading."

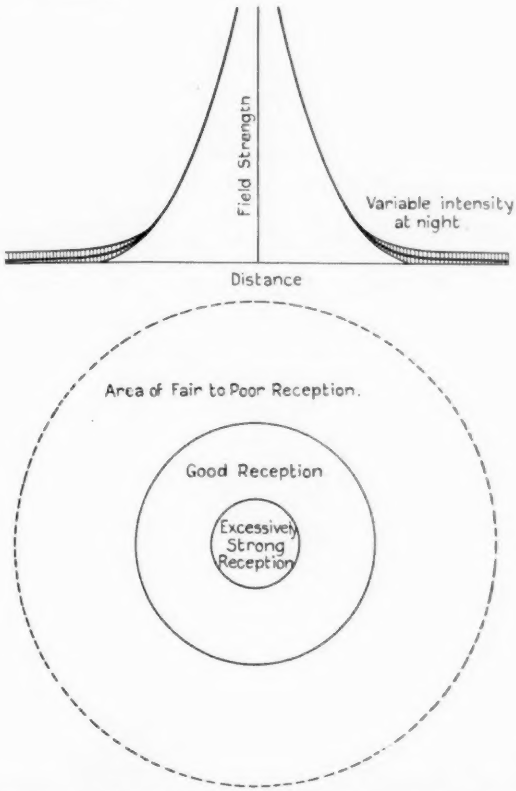


Fig. 1—The attenuation of broadcast waves in reference to the areas served

The evidence of recent researches, particularly those made at short wave-lengths, indicates that these fading variations are due to radiant energy which has left the earth's surface at the radio transmitter and has been reflected or refracted back to the earth's surface from a conducting stratum in the upper atmosphere. At broadcast frequencies the reflected wave component is observed at night but has not been noticed during the day. At locations close to the transmitting station the effect of the reflected component is negligible as compared with the strength of the directly transmitted waves. At

increasing distances the directly transmitted waves die away to very low values and the indirectly transmitted waves begin to show up and appear to become controlling at the longer distances. The fluctuations themselves appear to be due in part, if not entirely, to variations in the reflected waves themselves, resulting perhaps from fluctuations in the conditions of the upper atmosphere.

Thus, it seems clear that radio transmission involves wave components of two types: one which delivers directly to the receiving area immediately surrounding a broadcast station, a field capable of giving a reliable high grade reception; and another transmitted through the higher altitudes which permits distant reception but not with the reliability and freedom from interference required of high grade reproduction.

The effects which are actually realized in practice are indicated in a more quantitative manner by the curves of Fig. 2 which are plotted from some measurements made upon WEAF in New York and WCAP in Washington, D. C. They were made at locations in the New York and the Washington areas and at the intermediate points

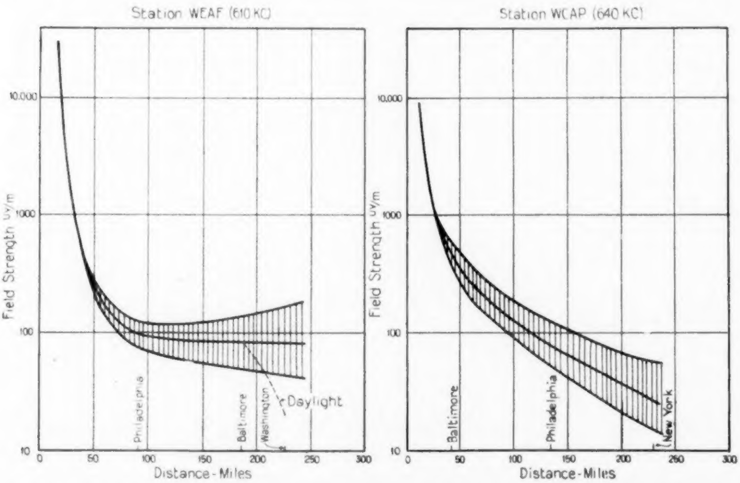


Fig. 2—Results of a few measurements upon the reduction in field strength with distance, including distances at which fading occurs

indicated on the curves. The measurements at each of these points are for one day only. They consisted in obtaining continuous graphic records of signal intensity during twenty-minute intervals out of each hour, one interval for each of the two stations. The period of time

covered for each set of measurements was that of from one hour before sundown to about three hours after sundown. The time of year was the latter part of May, 1926. The curves are plotted from an analysis of the records in terms of mean field strength. The range of variation due to fading is indicated by the shaded portions of the curves. The day and night fields were found to be roughly the same except for WEAF where there is a material drop in the daytime signal between Baltimore and Washington, shown in the WEAF curve.

Fading was observed to commence somewhere between 50 and 100 miles from the stations and the range of the fluctuations was found to increase up to the maximum distance observed. That the field of WEAF was found to be practically as strong at Baltimore as at Philadelphia is surprising. The data regarding this point are too meager, however, to enable any very definite conclusions to be drawn. The curves are useful principally in enabling the transition to be followed, in a more quantitative way than is done in Fig. 1, from field strengths capable of giving reliable reception, such, for example, as 10,000 μv./m. (microvolts per meter), to those which characterize the unreliable "distance" reception and are of the order of 100 μv./m.

A fact which is of importance to our understanding of these wave phenomena is that "fading," which ordinarily is noticed at distances of the order of 100 miles, may under some conditions become prominent at distances as short as 20 miles from the transmitting station. Such short-distance fading has been experienced in receiving WEAF in certain parts of Westchester County, New York.[3] It appears to be a case where unusually high attenuation, caused by the tall building area of Manhattan Island, has so greatly weakened the directly transmitted wave as to enable the effect of the indirect wave component to become pronounced at night.

In general, the attenuation suffered by the normal surface-transmitted waves varies over wide limits depending upon the terrain which is traversed. This is disclosed by the curves of Fig. 3, which show the drop in field strength with distance, for a 5 kw. station, for each of the following conditions:

a. No absorption, the inverse distance curve ($\alpha = 0$),

b. Sea water, for which the absorption is relatively small ($\alpha = 0.0015$),

c. Open country and suburban areas ($\alpha = 0.02$ to 0.03) as measured in the vicinity of New York and Washington, D. C.,

d. Congested urban areas ($\alpha = 0.04$ to 0.08) as measured for Manhattan Island.

[3] See "Some Studies in Radio Broadcast Transmission," by Ralph Bown, D. K. Martin and R. K. Potter; *Proceedings*, I. R. E., Feb., 1926.

The factor α will be recognized to be the absorption factor of the familiar Austen-Cohen empirical formula, which may be expressed as

$$e = 0.009 \frac{\sqrt{P}}{d} \epsilon^{-(\alpha d/\sqrt{\lambda})}$$

in which

P = radiated power in watts,
d = distance in kilometers,
λ = wave-length in kilometers,
α = absorption factor,
e = in volts per meter.

The first term represents the decrease in strength due merely to the spreading out of the waves; the second term, the decrease due to the

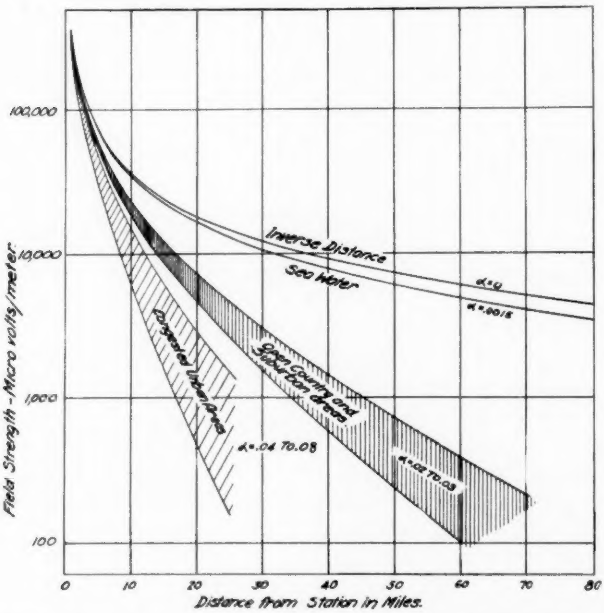


Fig. 3—Effect of the terrain in reducing the field strength of a broadcast transmitting station. 5 kilowatts in antenna, frequency 610 kilocycles, wave-length 492 meters

absorption of the wave energy by the imperfect conductivity of the earth's surface.

The curves given in Fig. 3 are derived from a considerable amount of data taken in the course of field strength surveys of the New York

City and Washington, D. C., areas. The results of some of the earlier of these surveys have already been published.[4]

Actual Distribution in New York City

Fig. 4 presents the results of a detailed survey of the field distribution effected over the New York metropolitan area by Station WEAF located at 463 West Street. The measurements upon which the plot is based were taken in the daytime during the summer of 1925. Measurements were taken at approximately one-mile intervals along each of a series of circular paths concentric with the station, the radii of which increased in steps of approximately five miles. The distribution was studied in even greater detail close to the station and in locations giving evidence of rapid change in field strength. Ferries were utilized to extend the measurements over bodies of water. Manhattan Island was circumscribed on water by measurements made upon a sight-seeing boat. The land measurements were made in all cases outside of buildings at ground level. In the built-up sections of the city they were taken in the middle of streets and street intersections, and in so far as possible in open places. The plot is based upon over 1000 measurements. While these measurements were taken over a considerable period of time, check measurements proved conditions to have remained quite stable and showed, in fact, little variation from measurements made the previous year. The type of measuring apparatus employed, together with certain of the results obtained in earlier surveys, has already been described.[5]

This plot is actually a simplification of a more detailed one. The number of contour lines has been limited to those of round figures for the sake of clarity. The line marked 10,000, for example, traces the locations at which that field strength was observed and beyond which lower values obtained.

This survey shows strikingly that the terrain over a city like New York is anything but uniform electrically; that the variations in the attenuation which the waves experience in different directions and from one area to another distort the distribution pattern from that which we might imagine from the familiar stone in the pool analogy. It is apparent that this simple analogy will have to be amended by conceiving the pool to be beset by various encumbrances causing high attenuations and reflections; and, in fact, also by the presence of

[4] "Distribution of Radio Waves from Broadcasting Stations over City Districts," by Ralph Bown and G. D. Gillett, published in the *Proceedings* of the Institute of Radio Engineers, August, 1924.

[5] See previous reference; also "Portable Receiving Sets for Measuring Field Strengths at Broadcasting Frequencies," by Axel G. Jensen; *Proceedings*, I. R. E., June, 1926.

surface ripples to represent the waves foreign to broadcasting which cause interference. Sight should not be lost, however, of the fact that the contour lines of Fig. 4 represent a two-dimensional section of

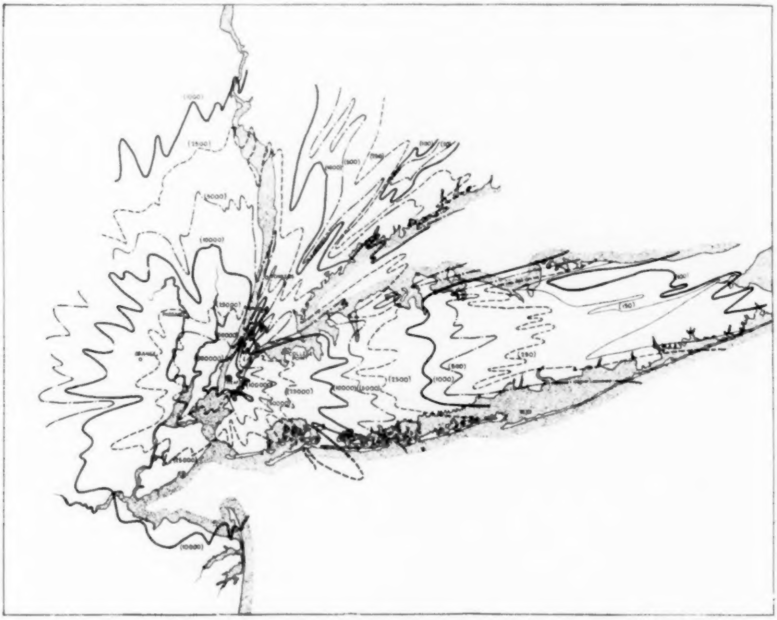


Fig. 4—Field strength contour map of distribution over the New York metropolitan area, effected by Station WEAF. 5 kilowatts in antenna, frequency 610 kilocycles, wave-length 492 meters.

a three-dimensional phenomenon. One should picture the contours as the intersections of the earth's surface with three-dimensional surfaces.

The fact previously referred to that the waves transmitted into Westchester County experience high attenuation is shown by the shape of the contour lines. The irregularity of the lines appears to be due to a splitting of the directly transmitted wave by the high building area and the filling in from the sides of wave energy transmitted along the two sides of the peninsula. Although the conditions in Westchester are quite stable during the daytime, they become unstable at night due, apparently, to the addition of the indirectly transmitted component reflected from above. An experimental study of this interference situation disclosed the fact that the bad quality

obtaining at night in certain parts of Westchester was due largely to a rapid frequency modulation of the broadcast transmitter. The frequency fluctuation of the transmitted band apparently caused the direct and indirect transmissions to slip in and out of phase rapidly. The use of a master oscillator control for insuring stability of frequency greatly improved matters, but evidence still remains of what might be called the normal night-time fading.

Another interesting effect which stands out in this map is the high attenuation of the wave-front transmitted over Long Island as compared with that which pursues the path of Long Island Sound and that of the ocean front to the south. The field over the eastern half of the island is contributed to by the water-transmitted waves from either side, giving rise to interference patterns similar to those in Westchester County.

A question which naturally arises is that of how strong a field, as measured in this way, is required for satisfactory reception. It is too early in the art to answer this question very definitely, for it depends first upon the standard of reception which is assumed, with respect to quality of reproduction and freedom from interference; and second upon the level of the interference. The interference, both static and man-made, varies widely with time and with location. It is therefore obviously impossible to give anything more than a very general interpretation of the absolute merit of field strength values. Observations made by a number of engineers over a period of several years in the New York City area, having in mind a high standard of quality and of freedom from interference, indicate the following:[6]

1. Field strengths of the order of 50,000 or 100,000 μv./m. appear to be about as strong as one should ordinarily desire. Fields much stronger than this impose a handicap upon those wishing to receive some other station.

2. Fields between 50,000 and 10,000 μv./m. represent a very desirable operating level, one which is ordinarily free from interference and which may be expected to give reliable year-round reception, except for occasional interference from nearby thunder storms.

3. From 10,000 to 1000 μv./m. the results may be said to run from good to fair and even poor at times.

4. Below 1000 μv./m. reception becomes distinctly unreliable and is generally poor in summer.

5. Fields as low as 100 μv./m. appear to be practically out of the picture as far as reliable, high quality entertainment is concerned.

[6] See also the paper by A. N. Goldsmith, "Reduction of Interference in Broadcast Reception," *Proceedings*, I. R. E., October, 1926.

Such fields, however, may be of some value for the dissemination of useful information such as market reports, where the value of the material is not dependent upon high quality reproduction.

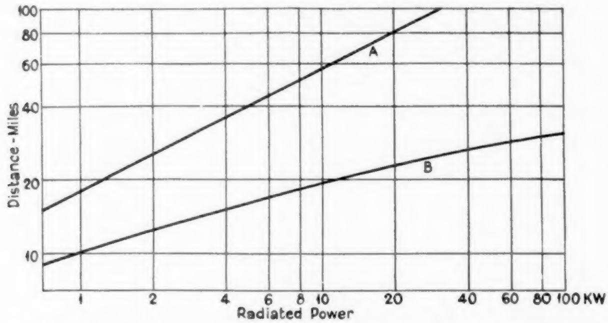


Fig. 5—Showing the increase in radiated power required to increase the range at which a field of 10,000 μv./m. is delivered. Curve *A* without absorption and curve *B* with absorption.

It is seen from the preceding three figures that a 5 kw. station may be expected to deliver a field of 10,000 microvolts some 10 to 20 miles away and a 1000 microvolt field not more than 50 miles. From this it will be evident that the reliable high quality program range of a 5 kw. station is measured in tens of miles rather than hundreds.

Higher Power Transmitting Stations Required

Rough though this interpretation of field strengths is, it indicates clearly the need which exists for the employment of higher transmitting powers. The range goes up with the increase of power disappointingly slowly. Even were no absorption present in the transmitting medium, the range in respect to overcoming interference would increase only as the square root of the increase in power. This is shown in the curve *A* of Fig. 5. It shows that a station which actually radiates five kw. of power would deliver a 10,000 μv./m. field at about 40 miles, a 20 kw. station the same field at distance 80 miles. Actually with absorption present the distances are less. This is shown by the curve *B* which gives the corresponding relations for the absorption observed for suburban and country terrain. To extend the 10,000 microvolt field from some 15 miles out to 30 miles would necessitate an increase in the radiated power from about five to 100 kw.

It is apparent from these relations that radio broadcasting is today underpowered; that the common 0.5 kw. station is entirely too small to serve large areas adequately, and that the more general use of

powers of the order of five kw. and even 50 kw. is decidedly in order. Such increases in power will be required if the broadcasting art is to be advanced to meet the higher standards of the future. The fact should be recognized that no greater interference between stations will be caused by the higher power levels, providing the increase in power is general among all stations. The interference difficulty arises in particular cases where one station suddenly makes a large increase and the others remain at their previous low power levels.

Mention should perhaps be made that the effect of raising the transmitter power in increasing the level of the *detected* signal is greater than would be inferred from the discussion above. This is because of the square-law action of the detector. In other words, the detector output reflects the increase in power of the carrier as well as the side band. In overcoming interference it is only the increase in side-band power which counts.

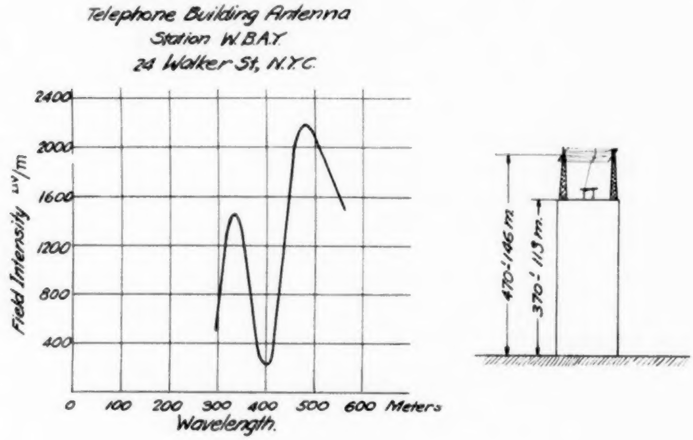
The ideal broadcast system from the transmission standpoint would be one in which the carrier is not transmitted from the sending station but is automatically supplied in the receiving sets themselves. This would save power, would reduce interference between stations and would reduce fading. It will be recalled that this system is being used to great advantage in the transatlantic radio telephone development. The practicability of employing it in broadcasting will depend upon receiving set development,—upon the economy with which carrier-generating receiving sets can be made and the ease with which the carrier frequency can be set and maintained with the necessary accuracy.

Transmitting Station on Tall Building

The location which naturally suggests itself for a broadcast station intended to serve a city is that of its center. Such a location might be expected to deliver the greatest strength of field to the greatest numbers because of the coincidence between high field strengths and high density of population. The other possibility, of course, is that of placing the station outside of the city, with the object of obtaining a better "get-away" condition, of covering a larger area and of laying down a more uniform, if less strong, field over the city itself. Instances of both of these types of locations readily come to mind. WEAF is a good example of a station located near the center of a large city. The results of a study which has been made upon the effect of moving the station to other possible locations are given below.

Before coming to this, however, there is another important factor to present and that is the effect of placing the transmitting station

upon a tall building. In locating a station near the center of a large city it is natural to select a tall building for the station site. This has been done for a number of stations in various cities. The operation of WEAF, when known as WBAY, was first attempted from the top of the 24 story long-distance telephone building located at 24 Walker



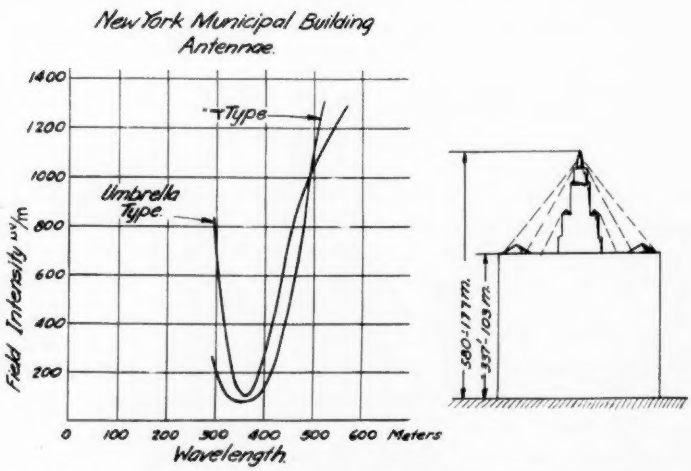


Fig. 6—The selective radiation characteristic of transmitting antennas on tall buildings

Street, New York City. It was found that with the limited wave-length range then open to broadcasting, radiation from the station was relatively poor. Measurements of the field strength delivered to a field laboratory located at Cliffwood, N. J. (on lower New York Bay), were made which gave the results shown in Fig. 6. The radiation was found to be sharply selective with respect to frequency, and to drop to a very low value at 400 meters. This happened to have been the wave-length assigned to the station at the time. When it became possible to shift the station to a longer wave-length, radiation was greatly improved, as indicated by the curve. The study made on this station was the first to disclose the fact that it is possible to have the building too high for the efficient radiation of certain frequencies.

As a result of this work it was possible to predict the probable occurrence of a similar effect in the case of a station which the City of New York desired to establish on the Municipal Building. Temporary antennas were erected and radiation from them measured at Cliffwood, N. J., using a transmitting oscillator of 100 watts. The results of these measurements are given in Fig. 6. The radiation was found to be a minimum in the vicinity of 360 meters, which was very nearly the wave-length which at that time was to have been assigned to this station. The establishment of the station at this location obviously could not be recommended until at a subsequent time when a longer wave-length was made available. The station is now operating on 526 meters, which is seen to be fairly well up on the radiation curve.

In both of these cases experiments were made with a number of different antenna arrangements and with different methods for driving the antenna and effecting the ground connection. None of the modifications, however, materially shifted the frequency of minimum radiation. This minimum occurs when approximately one quarter of the wave-length equals the height of the building. Measurements made upon buildings of lower heights have shown that for the usual broadcast wave-lengths heights of the order of 200 ft. are entirely satisfactory. The antenna of WEAF (which has been located for the past several years on the building of the Bell Telephone Laboratories, 463 West Street, New York), and that of WCAP, in Washington, are on buildings which put them at about this height above the street. They both have normal radiation characteristics.

Distribution from Suburban Locations

In order to determine the distribution over New York City which might be effected from locations outside of Manhattan Island, experi-

mental transmitting stations were established at each of several suburban locations. Use was made of an automobile truck equipped with a $\frac{1}{2}$ kw. broadcast transmitter and provided with a transportable

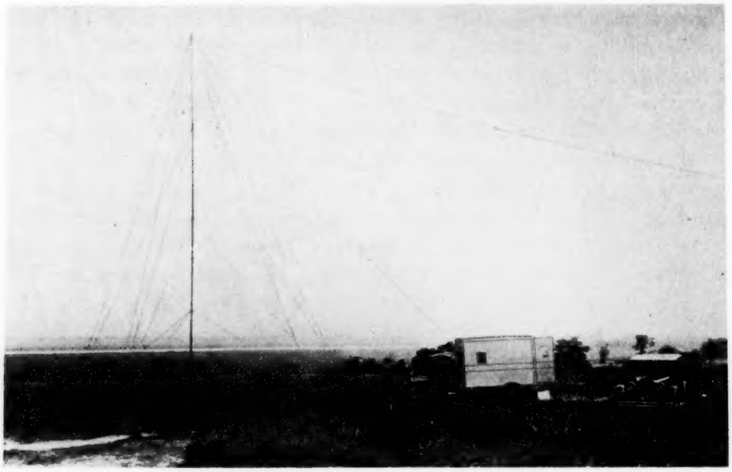


Fig. 7—Transportable field transmitting station

mast. The experimental transmitter as set up at Secaucus, New Jersey, is shown in Fig. 7. Measurements of the field strength delivered from each of the locations chosen were made over practically the entire metropolitan area. The results of these tests are given in Fig. 8, in comparison with those of transmission from the normal location of WEAF at West Street and from the earlier location at 24 Walker Street. The measured field strengths have been adjusted to correspond to the 5 kw. transmitter of the West Street station.

The smaller irregularities in the West Street curve as compared with the others are due to the greater detail with which these measurements were made. The curves should be compared merely with respect to their major contour characteristics. The inner contour line is for 50,000 μv./m. and the outer line for 10,000 μv./m. Actually, the measurements were made in sufficient detail to enable other contour lines to be drawn, but these have been omitted for the sake of simplicity.

The radiation from Secaucus will be seen to deliver a strong field to Manhattan Island, the most densely populated section, and, in general, to encompass the rest of the city quite well within the 10,000 μv. line. The irregularity in Queens County evidently represents the shadow cast by the tall building area on Manhattan Island.

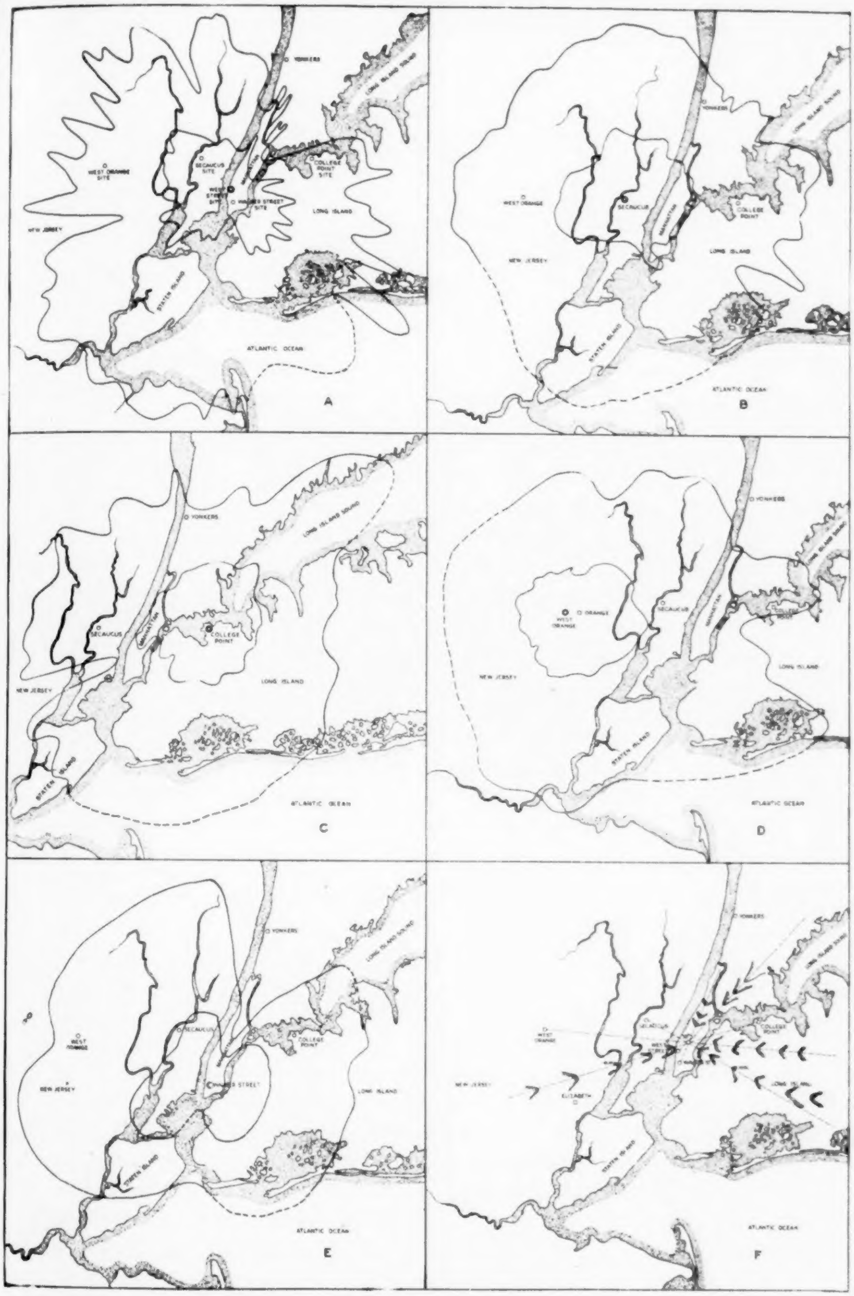


Fig. 8—Effect upon the field distribution of moving the transmitting station to suburban locations

A—463 West Street, New York City
B—Secaucus, N. J.
C—College Point, Long Island, N. Y.

D—West Orange, N. J.
E—24 Walker Street, New York City
F—Composite figure showing main shadows and center of obstacle

The distribution effected from the College Point location appears to be generally good. It does not cover the New Jersey suburbs as strongly as might be desired. The shadow cast by the Manhattan Island high buildings lies through Jersey City and lower Newark.

The distribution from the West Orange site appears to be somewhat less favorable. It is not sufficiently close in to deliver with moderate power a very strong field to the center of the population, nor is it sufficiently far out to avoid subjecting a considerable population in the immediate vicinity of the station to an excessive field were high power employed. The indent in the 10,000 μv. line in northern Queens is the shadow of the Manhattan buildings.

The distribution shown for the Walker Street location is seen to be generally similar to that of West Street. The curve presents a smoother appearance than the others because less data were taken in this one of the earlier surveys. The shadows cast to the north and to the south by the two areas of high buildings are prominent. Actually, a close examination of the contour lines reveals a noticeable angular displacement in the Westchester shadow as between Walker Street and West Street, Walker Street transmitting better up the Sound and West Street better up the Hudson. West Street turns out to be somewhat the better of the two.

The last diagram of the series brings together the shadows as determined from the several transmitting sites and shows that they project back to a common general center which locates at approximately 38th Street and Broadway, which corresponds quite well with the center of the up-town tall building area.

Relation between Wave Distribution and the Distribution of Listeners

The merit of a given distribution pattern obviously depends upon the relation which exists between it and the distribution of the receiving sets themselves. In order to study this relation more closely, the relative distribution of receiving sets was approximated by taking the distribution of residence and apartment house telephones in each of the central office districts of the metropolitan area, excluding the commercial telephones. It was assumed that the receiving set distribution is proportional to that of the telephones. For a given survey the field strength representative of each central office district is known. By assembling the figures for central office areas receiving like field strengths, and by doing this for the whole range of field strengths measured, an accumulative percentage curve may be derived which shows the percentage of the total number of receiving sets included within the contour lines of successively weaker fields.

Curves of this kind for each of the several surveys made are shown in Fig. 9. It will be seen that for field strengths of 10,000 μv./m. and better, the Secaucus and College Point transmitting sites include about 80 per cent of the receivers, that the West Street and West

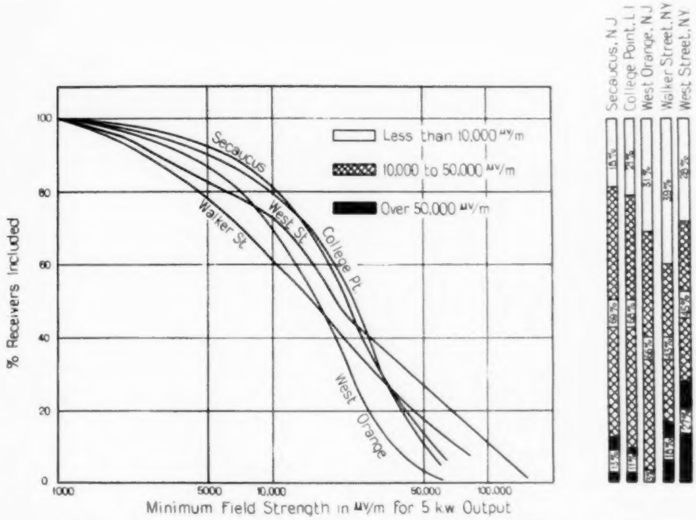


Fig. 9—Percentage of receiving sets in metropolitan area included within various strengths of field for each of the transmitting locations of Fig. 8

Orange sites include around 70 per cent and Walker Street about 60 per cent. These curves are further analyzed in the chart to the right of the figure to show in each case the proportion of the listeners which may expect to receive

a. less than 10,000 μv./m.,
b. between 10,000 and 50,000 μv./m.
c. over 50,000 μv./m.

It is seen from this that a location to the east or west of Manhattan Island would give a material improvement in uniformity of distribution as compared with a location on Manhattan Island. Had it been possible to include a station on Manhattan Island located farther north than is either West Street or Walker Street and included within the area of high steel buildings, it is probable that the corresponding curves for such a location would show the poorest distribution of the series.

The survey work described above did not go so far as to include a study of the distribution effected from a location well outside of the suburbs. The philosophy of such a location is, of course, that of attempting to encompass within the range of the station a widespread area and of so including the city within the area as to effect a more uniform distribution over it than is possible when transmitting from a location within the city. A theoretical study was made of the distribution to be effected from one such location in the general vicinity of Boonton, New Jersey, using attenuations obtained in the other surveys. Such a location would be somewhat similar to that of WJZ at Bound Brook, although the distance from Boonton to New York is less. The figures derived upon the basis of a 50 kw. broadcasting station are as follows:

Field Strength	Percentage of Receiving Sets in Metropolitan Area
Below 10,000 μv./m.	10 per cent
Between 10,000 and 50,000 μv./m.	79 per cent
More than 50,000 μv./m.	11 per cent

These figures show a good concentration in the most desirable field strength values. They should be discounted somewhat because they

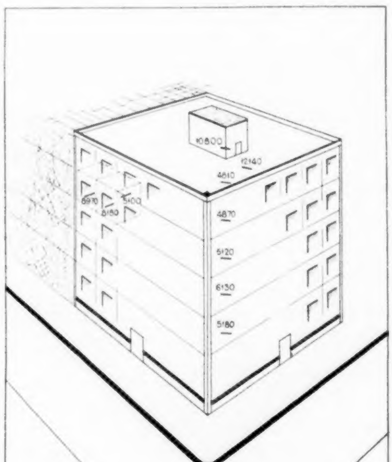


Fig. 10—The effect of non-steel apartment house building in shielding radio reception within it

are based upon symmetrical distribution and do not include the effect of irregularities, which an actual survey probably would reveal.

Receiving in Apartment Houses

The surveys described above disclose the field strength distribution as measured generally in the streets and open places. It does not disclose the details of field distribution in the immediate vicinity of a

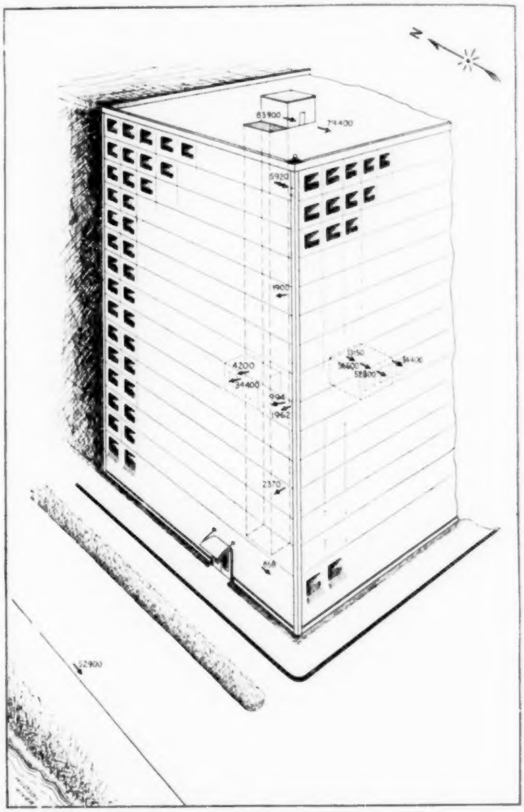


Fig. 11—The effect of steel structure apartment house building in shielding radio reception within it

receiver. Perhaps the most difficult situation is that of the large apartment house, particularly where it is desired to receive by means of an indoor antenna. Two effects are encountered: First, the reduction in signal strength by virtue of the shielding effect of the building;

and second, the existence of a relatively high noise level caused mainly by radio-frequency interference from electrical systems within the building.

The results of a few observations upon signal strength reduction within two buildings are presented in Figs. 10 and 11. Fig. 10, for a non-steel building, shows the field to be roughly halved. In the case of the steel structure building depicted in Fig. 11, the interior field is found to be reduced to as low as a few per cent of that outside the building. For outside rooms, the field strength near the window was found to be about eight times that further in the room. Such severe shielding effects obviously call for picking up the wave energy outside the building and conducting it to the receiving sets by wire circuits, preferably by shielded circuits, in order to protect against local interference.

Multi-Station Operation

The discussion given above has been directed chiefly to the relations which might be called internal to a single-channel radio broadcast system. Actually, of course, broadcasting involves the use of the common transmitting medium for a number of channels. This brings with it the problem of frequency selectivity and raises the question of the capabilities of the various types of radio receiving circuits.

In order to throw some light upon this important factor, measurements have been made upon a sample or two of each of a number of different types of radio receiving circuits. The measurements were made in the laboratory,[7] simulating as closely as possible the conditions under which the receiving sets would be used. The curves of Fig. 12 show the reduction which is to be expected in the detected audio-frequency current, were the receiving set tuned to a transmitting station on 900 kc., and the transmitting station then shifted in frequency by the amounts given along the abscissa. In this curve the reduction in current is indicated both as a ratio and in TU, which is a convenient way of indicating power ratios. The relation between TU and current ratio with a given impedance is indicated in the figure. Thus, for a carrier 40 kc. off from the one to which the set is tuned, the single-circuit, non-regenerative type of receiver

[7] The method consists in establishing a small laboratory transmitter and modulating it with a single-frequency tone. The receiving set is tuned to the modulated carrier signal as in practice. The gain or sensitivity of the receiver and its coupling with the transmitter are adjusted to produce normal load upon the detector tube. With the receiving set left at this adjustment the frequency of the radio transmitter is shifted each side of the original single frequency in 10 kc. steps throughout a range of 50 to 100 kc. For each of the offside frequencies the reduction caused in the detector output current is measured, this being an indication of the receiving set selectivity.

showed a cutoff of only 20 TU, corresponding to an audio-frequency current reduction to 0.1 that of the value at resonance. The curves will

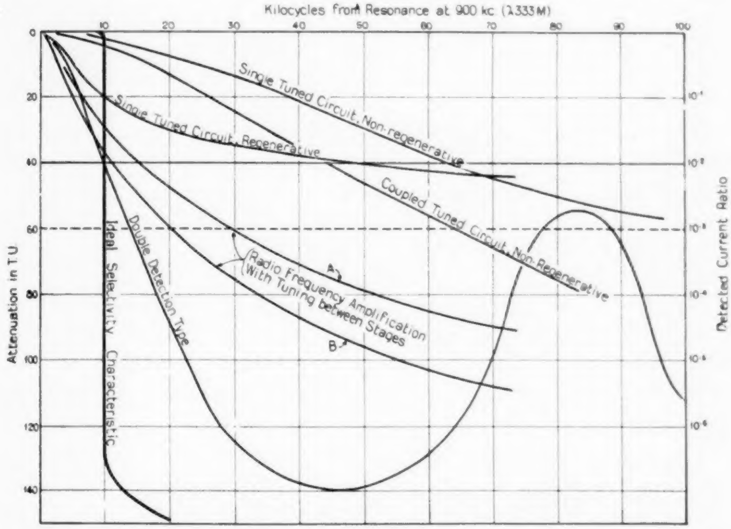


Fig. 12—Receiving set selectivity characteristics as measured from samples of receivers having different types of selective circuits

be seen to group themselves more or less into three classes in the order of their selectivity merit as follows:

1. The single-tuned circuit (non-regenerative and regenerative), and the combination of two tuned circuits coupled together.
2. Circuits employing radio-frequency amplification with tuned circuits between stages.
3. The double-detection or superheterodyne type of circuit.

The curve for the double-detection type of circuit shows a "comeback" which represents the familiar double-tuning effect. (Incidentally, the admittance of this particular set, which was not a commercial set, needs to be reduced by the use of more selectivity at the radio frequency.)

For comparison purposes there has been added to the figure the curve marked "ideal selectivity characteristic," in accordance with which the receiving set would pass without attenuation all frequencies up to 5000 or 10,000 cycles and would cut off abruptly all frequencies without this band. Attention is first called to the fact that the various circuits attenuate *within* the desired transmission band of

five or ten kilocycles. This means the higher frequency components of the side band will be reduced by the amounts indicated (after detection) with corresponding distortions of the reproduction. The distortion will be seen to be greater for the more highly selective sets. This follows from the nature of sharply tuned circuits. Selective circuits, capable of approximating the filter type of characteristic, are to be desired.

In comparing these selectivity characteristics, it is necessary to have in mind the amount of differentiation between the desired and the undesired signal which is necessary for the avoidance of interference. Each of the signals may be considered as fluctuating during the rendition of the program over a considerable range of volume which centers about some average value. The amount of differentiation required between the average values obviously depends upon the range of the fluctuations involved and upon the standard which is assumed with respect to freedom from interference. Experience with loud speaker reproduction indicates that ordinarily a level of the average of the undesired signal 40 TU lower than that of the desired signal, while not giving noticeable interference at times when the desired signal is strong, does permit the undesired signal to "show through" during times when the program rendition is weak. Reducing the undesired signal to 60 TU below the desired signal prevents this interference for the volume ranges which are now commonly transmitted. If the future art brings with it the requirement of following greater swings of volume, a further reduction in the undesired signal may be necessary. The value of 60 TU has been dotted in across the chart of Fig. 12, in order to show readily the frequency separation at which the different selective circuits give this attenuation of the undesired signal. This is upon the basis that the field strengths of the two signals are equal. Inequalities in field strength require that the 60 TU value be increased or decreased by the amount of the inequality as measured in TU.

The frequency interval which has been recommended by the National Radio Conferences for stations in the same zone is 50 kc. It is evident from the curves that sets equipped with the simpler types of tuned circuits will be subject to some interference between stations thus separated even if the receiver is so favorably situated as to receive equal field strengths from the desired and undesired stations. The selectivity of the other types of receiving circuits is seen to be sufficient to avoid interference under these conditions and allow some margin for overcoming inequalities between the fields. Such inequality becomes great where the attempt is made to receive distant

stations through the effect of local stations. Assume, for example, that the listener receives 50,000 microvolts from a local station and 500 microvolts from a distant station to which he desires to listen. There exists a 100 to 1 or 80 TU disadvantage to be overcome. When added to the 60 TU needed for crosstalk clearance, the total selectivity requirement, as measured in terms of detected audio current, becomes 140 TU, or a current reduction of the order of 10,000,000 to 1. The need for a high degree of selectivity is therefore apparent. The impracticability of receiving distant stations removed in frequency from local stations by any such narrow margin as 10 kc. is also obvious.

The effect of receiving set selectivity in increasing the area over which a station may be received without interference from a second station is illustrated in Fig. 13. The two stations are assumed to be of equal powers so that they deliver equal field strengths to receiving stations along a line midway between them. Receiving sets so located are required to have an amount of selectivity called for by the crosstalk margin itself, say 60 TU. On the desired-station side of this line the selectivity may be less; this is the region where poorly selective receivers can be employed. On the undesired-station side of the center line the selectivity requirements are greater. The non-interference area is pushed up closer and closer to the undesired station as the receiving set selectivity is improved, as is indicated by areas *A* and *B* of the figure. For example, assume that the selectivity of the receiving set is such as to give a 100 TU cutoff of an undesired station, offset by 50 kilocycles. Sixty TU of this would be required were the two signals of equal strength, so that 40 TU measures the difference by which the undesired signal may be greater than the desired one. The increased area of reception made possible by this additional 40 TU is indicated by that portion of the lower figure which is to the right of the center line and outside of the area *A*. Within the area *A* interference would be suffered. This interference area may be diminished by the use of still greater selectivity. The addition of another 20 TU of selectivity (again as measured in terms of detected audio-frequency current) would reduce the interference area to that within the small area at *B*. The extent to which the selectivity requirement of the receiving set is determined by its location, therefore, is apparent. The conditions which obtain in multi-station areas, such as New York City and Chicago, obviously call for a general use of high selectivity sets.

In locating a new transmitting station it should be possible from a knowledge of the relative field strengths of other stations in the vicinity to predict approximately what the interference area will be

for the different types of receiving sets. In this connection there should be recognized the advantage from the interference standpoint which exists in grouping together the broadcast transmitting stations as far as possible in one location, and in equalizing their powers. Such

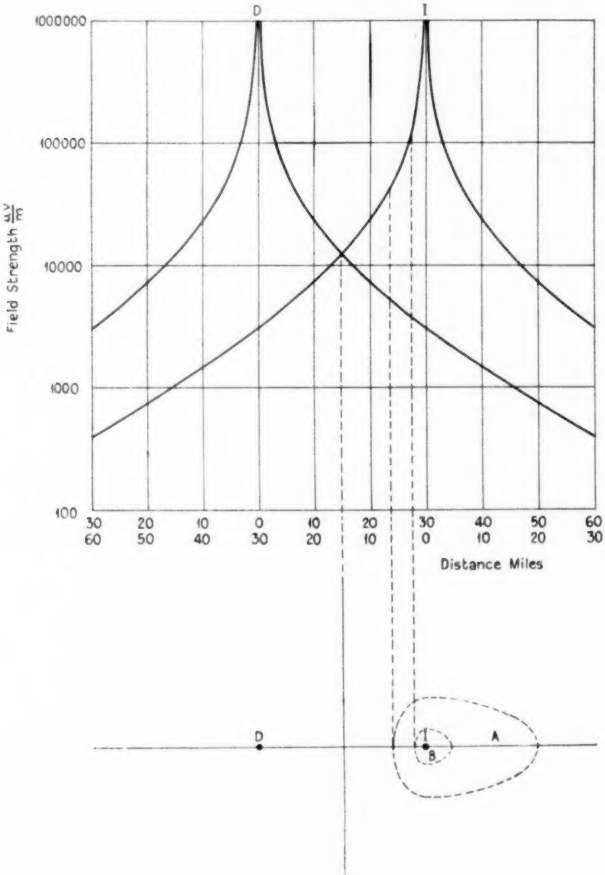


Fig. 13—Showing the greater area over which the more highly selective receiving sets may receive a desired station D and exclude an interfering station I

grouping and equalizing would enable the receivers to obtain substantially equal fields from all of the stations and would minimize the selectivity which they are required to possess. While it is im-

practicable to accomplish this result completely, it is hoped that a better understanding of the interference problem as here outlined and of the mutual advantage to be gained in reducing interference will lead naturally to a better coordination of radio broadcast stations.

Acknowledgment

The data and analyses presented in this paper are the result of the cooperative effort of a number of engineers in both the American Telephone and Telegraph Company and the Bell Telephone Laboratories, Inc. In assembling and presenting the material the writer is acting merely as the spokesman for these development groups. He wishes to acknowledge his indebtedness to his colleagues, particularly to those who have directly participated in the survey work described and assisted in the preparation of the paper, namely, to Messrs. D. K. Martin, R. K. Potter, G. D. Gillett and H. B. Coxhead, and to Messrs. S. E. Anderson and O. O. Ceccarini of the Bell Telephone Laboratories to whom is due the measurement work upon the radio receiving sets.

A Shielded Bridge for Inductive Impedance Measurements at Speech and Carrier Frequencies [1]

By W. J. SHACKELTON

SYNOPSIS: A shielded, a-c., inductance bridge adapted to the measurement of inductive impedances at frequencies up to 50,000 cycles is described. The bridge comprises a balancing unit and associated standards of inductance and resistance. The balancing unit has resistance ratio arms specially constructed to meet the requirements imposed by the above frequency range. The reference standard makes use of inductance coils of a new type, their cores being of magnetic instead of non-magnetic material as is usually the case. The use of such cores results in coils that are smaller and hence better adapted to assembly in a multiple shielded standard.

The bridge is completely shielded so as to eliminate, to a high degree, errors due to parasitic capacitance currents. The shielding is also arranged so as to permit the correct measurement of either "grounded" or "balanced-to-ground" impedances. A series of diagrams is shown for the purpose of indicating the function of each part of the shielding system.

Equations expressing the errors resulting from any small residual capacitance unbalances in the resultant bridge network are given and calculations made of the balances required for the desired degree of measurement precision. Test data are presented illustrating a method of experimentally checking the residual shunt and series balances from which it is concluded that the bridge is capable of comparing two equal inductive impedances of large phase angle with an accuracy at the maximum frequency of 0.02 per cent for inductance and 1.0 per cent for resistance.

INTRODUCTION

THE limitations of the ordinary unshielded bridge network as a means of making precise a-c. measurements at speech frequencies were early recognized by telephone engineers. The solution of cross-talk problems arising in connection with the use of cable circuits was found to require an exact knowledge of the capacitive balances existing between such circuits at speech frequencies. For the ready and accurate determination of the capacitances defining these balances, together with their associated conductance values, G. A. Campbell devised the "shielded balance." [2] This is a bridge network having its parts individually and collectively shielded so as to define exactly the mutual electrostatic reaction of each with respect to all other parts of the electrical system affecting the balance condition.

As a means of more completely treating the cross-talk problems of cable circuits, Campbell conceived also the very valuable idea of "direct capacity" as distinguished from the "ground" and "mutual capacities" in use up to that time.[3] The shielded balance was found

[1] Presented at the New York Regional Meeting of the A. I. E. E., New York, N. Y., Nov. 11–12, 1926.

[2] G. A. Campbell: "The Shielded Balance," *Electrical World* and *Engineer*, April 2, 1904, p. 647.

[3] G. A. Campbell: "Measurement of Direct Capacities," BELL SYSTEM TECHNICAL JOURNAL, July, 1922, p. 18.

to be especially adapted to the precise measurement of direct capacities employing the substitution method devised by E. H. Colpitts.[4] Shortly thereafter, with the advent of loading for telephone lines, the same principles of shielding were extended to apply to bridge networks specially arranged for the measurement of the speech-frequency inductance and effective resistance of loading coils. As the successful commercial application of loading required the manufacture of these coils in large numbers to precise requirements, it was quite essential that testing means be available permitting a relatively unskilled tester to determine quickly whether the proper adjustment of the coils had been made. For this purpose the shielded balance has proved to be extremely valuable. More recently the employment of frequencies up to 50,000 cycles for carrier telephone and telegraph purposes has led to the need for correspondingly precise measurements at these higher frequencies. In this field the advantages of the shielded bridge are so great as to make it almost indispensable.

While the fundamental principles of the shielded balance are essentially the same for all impedance measurements, the practical application of shielding to any concrete bridge problem may vary according to the kind and range of impedances to be tested, the frequency range to be covered, and the precision required. It also presents special problems in the design and construction of several of the circuit elements. This paper describes a particular form of shielded bridge which has been developed to meet the conditions commonly encountered in the measurement of inductance at speech and carrier frequencies. The facts leading to the detailed construction are discussed and some experimental data given to illustrate the performance of the bridge.

General Features

A simple schematic diagram of the bridge circuit is shown in Fig. 1. To avoid confusion, no shielding is shown in this diagram. As will be noted, there are provided two equal non-inductive resistance ratio arms, an adjustable standard of self-inductance, an adjustable resistance standard, a thermocouple milliammeter, two transformers and two adjustable air condensers. Physically, this apparatus is grouped into three separate units, one comprising the standards of inductance, one the resistance standard, and the third, the remaining parts of the circuit. The last assembly constitutes what may be considered the balance element of the system, by means of which the unknown and standard impedances are compared. Figs. 2 and 3 show the arrange-

[4] See Note 3.

ment of the parts in this unit. Fig. 4 illustrates the appearance of the standard inductance unit and Fig. 5 shows how the units are associated when a test is being made. The thermocouple milliammeter indicates the total effective test current applied to the bridge and forms a

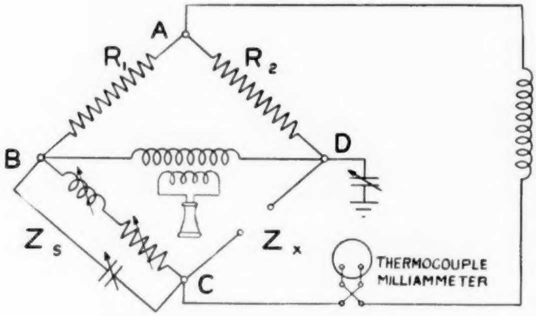


Fig. 1—Schematic diagram of bridge circuit

means of determining when this current has been adjusted to the desired value.

In operation, the air condensers are first adjusted to produce an initial or zero balance of the residual electrostatic capacitances of the apparatus. Aside from the initial balancing, the operation of the bridge follows the usual practise; that is, the standards of inductance

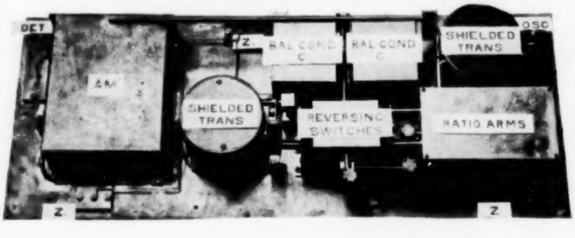


Fig. 2—Balance element of shielded bridge. Rear view of panel removed from case

and resistance are alternately adjusted until the balance detector indicates a condition of zero potential difference at every instant between the bridge points to which it is connected. The inductance and resistance values as indicated in the standard arm are then equal (within the precision limits of the bridge) to the corresponding constants of the unknown impedance.

Purpose of Shielding

The principal difficulties in attaining a satisfactory degree of precision in inductance measurements at relatively high frequencies by means of unshielded bridges are those due to the presence of residual or stray admittances existing between the bridge parts or from them to ground. All these parts have quite appreciable surface dimensions

Fig. 3—Balance element of shielded bridge. Front view of panel and case

and when exposed at the usual separations to each other or to ground, have corresponding direct and grounded admittances. Leads to the source of testing current and to the balance detector also introduce rather large admittances. In a bridge intended for rapid operation, the parts subject to manipulation must be arranged compactly and

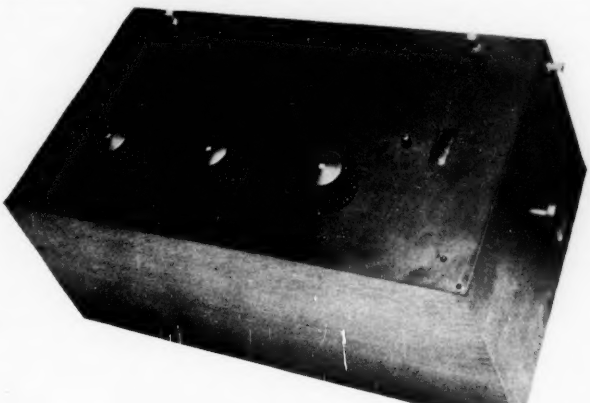


Fig. 4—Inductance standard

conveniently to the operator. This makes it impracticable to isolate them sufficiently to make the admittance values between these parts

and between them and ground (the operator being considered to be at ground potential) negligibly small.

To make the matter more concrete, there is shown in Fig. 6 a

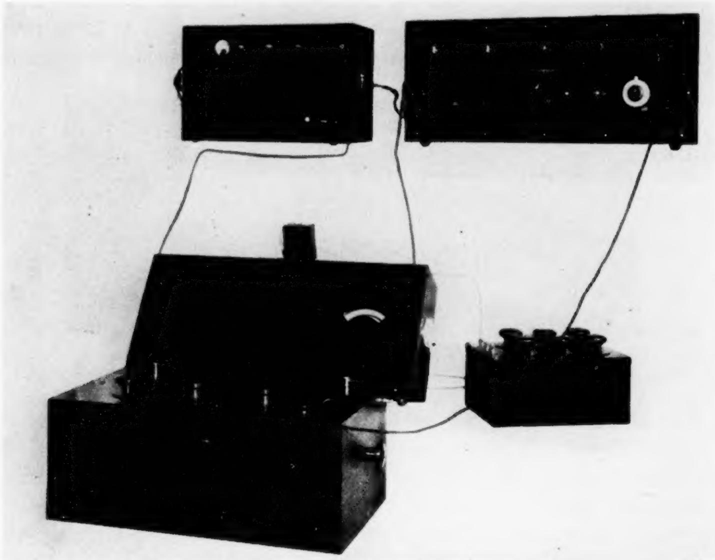


Fig. 5—Shielded bridge connected to vacuum tube oscillator and heterodyne detector

schematic diagram with possible positions of some of the more important of these admittances indicated as at C_1, C_2, etc. (With some exceptions, the capacitance components of these stray admittances substantially determine their full effect. In the diagrams and discussion, therefore, the conductance component will be neglected except where its effect is significantly large.) The capacitances between the two ratio arm coils, R_1 and R_2, and from each to ground, are shown as being uniformly distributed along the length of the coils symmetrically with respect to each other. If this symmetry is perfect these capacitances do not affect the bridge balance. In practise, however, they will only be approximately so, with the result that the two arms will be somewhat unbalanced to alternating currents, the effect of the unbalance increasing with the frequency. While the ratio arm capacitances can be made fairly small, others such as those indicated at C_1, C_2, C_3, and C_4 will commonly be much larger and hence of greater effect. Capacitances C_1 and C_2 are frequently comparatively large

due to the use of long distributing wires, encased in grounded conduit, for supplying the testing current. C_3 may consist chiefly of the ground capacitance of the outer layer of the detector coil winding and C_4 that of dead-end coils of the reference standard, Z_S.

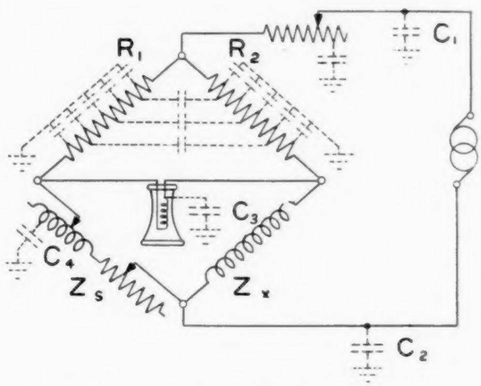


Fig. 6—Bridge circuit with stray admittances

Some of the currents flowing along the paths provided by these capacitances will complete their circuits external to the bridge network proper and will not affect the balance; for example, that through capacitances C_1 and C_2 in series. Other currents, however, will flow unsymmetrically through parts of the bridge circuit; for instance, that through C_1 and C_3 in series and the arm Z_X; also, that through C_2 and C_4 in series, returning through the ratio arm R_1. These latter currents and others of the same sort affect the potential distribution of the bridge and hence the values of the impedances required for balance. Certain of these capacitance currents in the bridge network tend to neutralize or balance the effects of others; for example, that through the arm Z_X due to the series action of capacitances C_1 and C_3 has a balancing effect with respect to that through C_1 and C_4 and the arm Z_S and would be without reaction on the bridge balance if capacitances C_3 and C_4 were exactly symmetrical with respect to the two detector terminals. Such balancing, however, is accidental in nature, seldom satisfactorily complete and, in part, not constant. Even were it made approximately complete for a particular arrangement, the substitution of another detector or the use of another source of testing current would probably destroy the balance. Variable effects would always be present; for example, those due to the changing position of the operator relative to the parts of the circuit or the

effects of parallel loads on the supply generator. The distribution and value of the ratio arm ground capacitances described above are functions of the bridge surroundings; hence they are also subject to change if the bridge is moved from place to place. In the bridge being described, however, the shielding used affords a means of definitely fixing and controlling the various inter-circuit capacitances. Consequently, such variations cannot take place, balances between the resultant capacitance currents can be made as desired, and the bridge measurements are satisfactorily precise.

Shielding System Used

It is felt that the merits of the particular shielding system adopted for this bridge can best be brought out by showing, step by step, the reasons for using each of its elements.

The first step is to simplify, for further treatment, the initial residual capacitance network of the unshielded circuit. This is done by providing individual shields for each part of the circuit that it is desired to have function as an independent unit. Such shields can be connected to one of the terminals of the part enclosed and thus there is substituted, from the standpoint of terminal-to-terminal characteristics, a definite and invariable condition in place of that which was previously a function of the relation of the part to its surroundings. For example, as shown in Fig. 7, shields would be placed

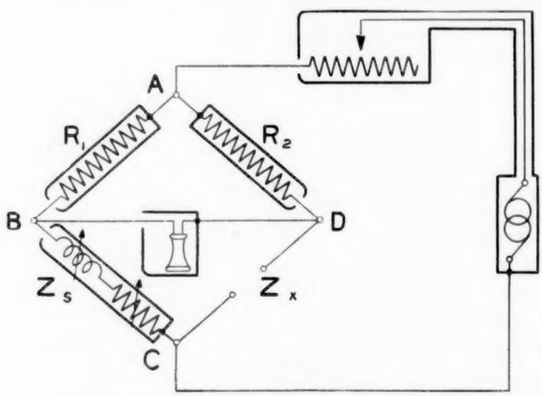


Fig. 7—Bridge circuit with local shields

around the resistance coils forming the ratio arms R_1 and R_2 and connected to the junction point A of the system, one enclosing the elements of the standard impedance Z_S and another around the source

of testing current and connected at *C*; likewise, one around the detector is connected at *D*. It will readily be seen that these shields localize the effects of the various capacitance currents. Those circulating within the shields have, of course, no effect exterior to the shields, while those flowing between the various shields directly or by way of ground enter and leave the bridge system at definite points. By themselves, these shields do little good but they are necessary in order to make the next step, the balancing of the capacitances, practicable.

Generally it will not be found convenient to shield the current supply apparatus, especially if this is a power-driven generator. Also, to promote greater flexibility in respect to testing with a wide range of frequencies, it will often be desirable to substitute one source of current for another and likewise one detector for another. The shielding of this apparatus should therefore be reduced to a minimum. This is readily effected by making both the supply and detector branches of the bridge one of the windings of a transformer. This winding can be electrostatically shielded without affecting its transformer action and then any desirable source of current supply or any type of detector can be magnetically coupled with it.[5] Introducing this change the circuit becomes as shown in Fig. 8. The

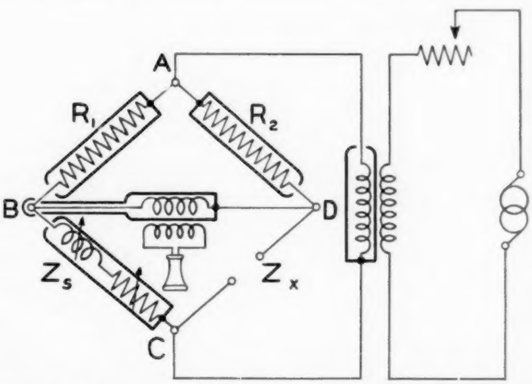


Fig. 8—Bridge circuit with shielded transformers and local shields

capacitances of the various shields to ground being still variable, the next step to correct this condition would simply be to add a ground shield around each. At this point, however, it becomes necessary to consider the ground admittance relations of the impedances to be tested.

[5] U. S. Patent No. 792248, June 13, 1905.

In general, the unknown impedance will have capacitances to ground and the effect of these will be properly included in the measurement only when certain conditions as determined by the nature of the apparatus are fulfilled. From this standpoint the impedances usually encountered are of three general classes: (1) Those having ground admittances negligibly small in comparison with the direct terminal-to-terminal admittance; (2) those having appreciably large admittances to ground approximately balanced with respect to the two test terminals; (3) those having one terminal directly grounded, the other having an appreciably large ground admittance.

In measuring apparatus of the first type it is evident that since in connecting it to the bridge circuit no additional ground admittances are introduced, the balance between those previously existing can be made without reference to the test impedance. The connection of an impedance of either of the other types will, however, introduce additional ground admittances into the bridge system, which, unless precautions have been taken, may cause the result to be something other than that which is wanted. In general, the desired test is that which gives the effective impedance applying to the apparatus as it is used. In the case of impedances having balanced admittances to ground, this is the effective value of the direct, terminal-to-terminal impedance as modified by the effect of the two ground admittances acting simply in series with each other. This condition is obtained when equal currents flow in each of these admittances, or, what is equivalent, when the electrical potentials of the terminals are balanced with respect to ground potential. To obtain this condition, when the impedance is being tested, the bridge terminals to which it is connected must likewise be balanced with respect to ground potential; that is, ground potential must be at the midpoint of the unknown impedance arm. If the only admittances to ground of the bridge system are those of the junction points (as is the case in Fig. 8), the potentials of these points with respect to ground are entirely determined by these admittances. To make any two points, such as the terminals of the unknown arm, have equal potentials to ground, it is sufficient to concentrate all of the ground admittances to these or other equipotential points and then balance the admittances from each. Referring to Fig. 9, if the testing current is applied at the points A and C, this condition is realized as shown by concentrating all ground capacitances at junction points B, C and D, and making the sum of the capacitances of junction points B and D equal to that of junction point C. This follows from the fact that when the bridge is balanced the junctions B and D are equipotential points. The mid-point of arm CD is now

at ground potential. If, however, the testing current is applied at the points *B* and *D*, the equipotential points are the junctions *A* and *C*, the sum of whose ground admittances would then be made equal

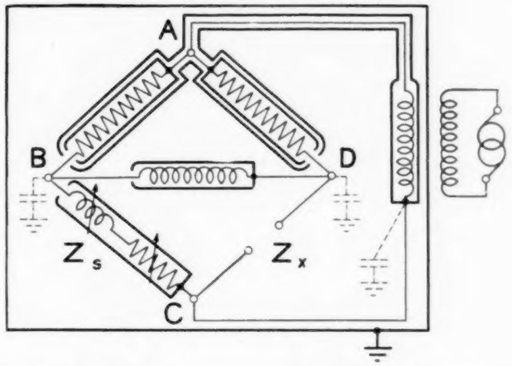


Fig. 9—Shielded bridge circuit showing location of ground admittances

to that of *D* and the arm *CD* again balanced with respect to ground potential. In this case there must be no ground admittance from junction *B*. To permit of testing under both conditions, point *A* and all connected conductors are protected with a shield which is then connected to the point *C*. Point *B* is likewise enclosed by a shield connected to point *D*. These two main shields then represent the junction points *C* and *D* of the bridge and are fixed with respect to capacitance to ground by a ground shield which may be common to the two.

There now exist, external to the local shields, direct capacitances only between points *A* and *C* and between *B* and *D* (which do not affect the bridge balance), and from points *C* and *D* to ground. These latter do, of course, affect the balance. Two courses are open. Their effective resultant value shunting the arm *CD* can be determined and allowed for by calculation. Such calculations would involve a considerable amount of labor, however, and can be avoided very simply by providing in the opposite arm an exactly equal shunt capacitance. To permit adjusting the ground capacitances of points *C* and *D*, an adjustable condenser is connected to ground from the point having the lower value. With the apparatus connected as shown this is usually point *D*. The shielded system then becomes as shown in Fig. 10.

When impedances, which in actual service are grounded at one terminal, are to be tested, the matter is much simpler. Then it is

necessary only to definitely ground one of the bridge terminals to which the impedance is connected and establish the proper initial capacitance balance of the bridge for this condition. This is readily done by grounding junction point *C* and adjusting the capacitance from *B* to *C* to equal the ground capacitance of *D*. The shielding system may remain the same as in Fig. 10.

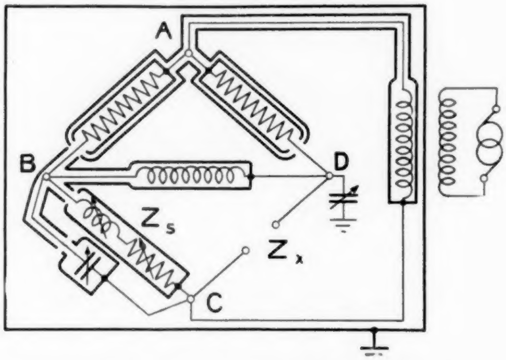


Fig. 10—Shielded bridge circuit with balancing condensers

In the case of the bridge being described it was desired to have a means of verifying by reversal the degree of balance of the ratio arms and also that of the impedance arms. The bridge is therefore equipped with reversing switches for this purpose. Due to the appreciable effect produced by a relatively small capacitance unbalance arising from factors present only when the arms are in circuit, it is quite important to be able to do this when a high degree of accuracy is desired. To effect the proper reversal, however, certain conditions must be definitely maintained. In reversing the impedance arms none of the inherent bridge admittances should be disturbed; that is, only the unknown impedance and the standard as read should be transferred. On the other hand, in reversing the ratio arms not only should the resistance element of these arms be transferred but also all associated shunt admittances. Moreover, in transferring these admittances they must be absolutely unchanged. A further requirement is that the ratio arm reversal must not occasion the shifting of any capacitances shunting the impedance arms. To accomplish these objects a suitable arrangement of shielded switches was worked out and added to the circuit of Fig. 10, the result being as shown in Fig. 11. In this arrangement all capacitances between the various parts of the switches which are subject to change due to physical movement of

the switch parts are either short-circuited or connected across opposite bridge points and hence do not affect the bridge balance. The small capacitance C_R between the switch shield and that of the ratio coil R_2 shunts this coil and is not carried with it on reversal. For this reason a corresponding capacitance C_R', shunting the coil R_1, is pro-

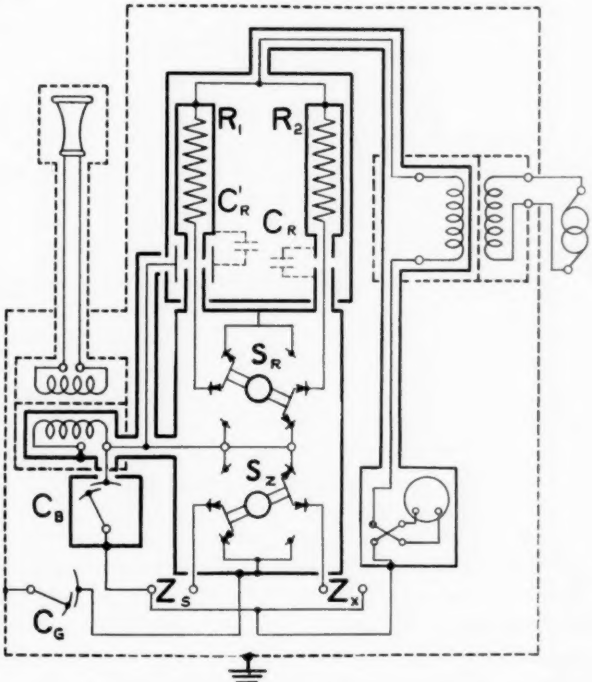


Fig. 11—Complete circuit diagram of balance unit with shielding

vided and connected to the opposite point of the switch. This is adjusted by test to equal the value of C_R. The diagram of Fig. 11 represents completely the circuit and shielding used for the balance unit of the bridge.

While, from the standpoint of the bridge balance alone, the parts comprising the standard impedance can be shielded with a local and ground shield as shown in Fig. 9, unless the standard has a very limited range, the resulting calibration is exceedingly laborious to make and use. To reduce calibration difficulties, additional shields can be used; this, of course, increasing the cost of construction. In arriving at the

proper compromise between these conflicting factors the size and impedance value of the part to be shielded must be considered. This question will therefore be taken up in more detail in the following section.

Construction

The circuit and shielding features discussed so far are of general application to impedance measurements without restriction as to the particular range of values to be tested or frequencies to be used. The physical construction is, however, dependent upon these factors. As initially stated, the bridge is intended for the measurement of audio and carrier frequency inductances. By this is meant all apparatus having reactance values nearly equal to the respective impedance values. For the purpose of the present discussion, such inductances will be more exactly defined as those having ratios of reactance to resistance of not less than 10 (minimum phase angle of 84 deg., 20 min.). The difference between the reactance and the impedance of any such inductance does not exceed ½ per cent. The impedance values range from about 100 to 10,000 ohms and testing frequencies from 500 to 50,000 cycles.

On the basis of these conditions, the following construction was developed and is used for this bridge.

Ratio Arms. It is desirable from the standpoint of sensitivity of balance to have the ratio arm impedances of approximately the same value as those of the other two arms. Considering the range of impedances to be covered and giving due weight to the values which are of most importance in telephone circuits, a ratio arm resistance of 1000 ohms was selected. The problem then was to construct two 1000-ohm resistances, balanced both as to effective resistance and effective inductance for a frequency range from 500 to 50,000 cycles when subjected to the usual temperature and humidity variations.

Curtis and Grover have discussed the factors affecting the characteristics of a-c. resistances and have suggested forms suitable for general use at frequencies up to 3000 cycles.[6] A 1000-ohm resistance, constructed according to their specifications, is made by winding with a 1/10-mm. diameter, double-silk-covered manganin resistance wire, five 200-ohm bifilar sections on a 1-in. spool of insulating material. These sections are spaced about three mm. apart on the spool and are connected in series to form the 1000-ohm coil. Such a coil, when shellacked, baked and coated with paraffin, was found to be substantially constant in resistance and to have constant phase-angle

[6] H. L. Curtis and F. W. Grover: "Resistance Coils for Alternating Current Work," *Bulletin of the Bureau of Standards,* Vol. 8, No. 3.

effects equivalent to shunting capacitances of the order of 10 to 15 mmf. for all frequencies up to 3000 cycles. Since individual coils made according to this method may differ in their effective capacitances by as much as five mmf., some adjustment of these capacitances (as well as of the resistance) is required in order to make them suitable for use as the required ratio arms. Assuming that this is done by adding to the coil having the lower value a small capacitance of suitable constancy, it may be concluded that two coils so balanced would be suitable for use at frequencies up to 3000 cycles.

In arriving at the requirements for the more extended frequency range of this bridge, the necessary phase-angle balance was first considered. Designating by L_X and R_X the inductance and effective resistance of the impedance being tested, and by L_S and R_S, the corresponding components of the standard impedance required to balance it in a bridge circuit having ratio arms of exactly equal resistances R but shunted by slightly different capacitances, C_1 and C_2, and assuming that the quantities are such that $\omega^2R^2C_1^2$ and $\omega^2R^2C_2^2$ are small in comparison with unity, the equation for balance is

$$(R_X + jwL_X)(R - jwC_1R^2) = (R_S + jwL_S)(R - jwC_2R^2),$$

which reduces to

$$R_X = R_S + w^2R(C_2L_S - C_1L_X) \tag{1}$$

and

$$L_X = L_S - R(C_2R_S - C_1R_X). \tag{2}$$

Neglecting second order effects, these can be written

$$R_X = R_S + w^2RL_X(C_2 - C_1) \tag{3}$$

and

$$L_X = L_S - RR_X(C_2 - C_1). \tag{4}$$

If the readings R_S and L_S are taken as the values of the unknown resistance and inductance, respectively, it is evident that errors as given by the last terms of these equations will be present. The percentage errors in the two cases are as follows:

$$\begin{aligned}\Delta R_X\,(\%) &= 100\omega^2R(C_2 - C_1)\frac{L_X}{R_X} \\ &= 100\omega R(C_2 - C_1)\tan\theta, \qquad (5)\end{aligned}$$

$$\Delta L_X\,(\%) = 100R(C_2 - C_1)\frac{R_X}{L_X}\,. \tag{6}$$

For a given capacitance unbalance of the ratio arms, it is seen that the error in inductance is inversely proportional to the time constant L/R of the impedance arm and is independent of the frequency, while the error in resistance is proportional to the frequency and to the ratio of reactance to resistance, that is, to the tangent of the phase angle. The inductance error is, therefore, maximum for the minimum time constant apparatus to be tested. Within the range previously mentioned this occurs when an impedance having the minimum reactance to resistance ratio of 10 is being measured at the minimum frequency of 500 cycles. Under this condition R/L has a value of $(2\pi \times 500)/10$ or approximately 300. The corresponding percentage error in inductance per micro-microfarad of capacitance unbalance is then $300 \times 1000 \times 10^{-10} = 3 \times 10^{-5}$ or 0.00003 per cent. Evidently a very considerable unbalance can be tolerated. In the case of the resistance component, the error is maximum when an unknown impedance having the maximum reactance to resistance ratio is being tested at the maximum frequency. A reactance to resistance ratio of 300 is very rarely exceeded. For this value, the error per micro-microfarad unbalance at a frequency of 50,000 cycles amounts to about 9.5 per cent. Hence, to limit the error from this source to the order of 1 per cent requires a balance of about 0.1 micro-microfarad. It will be appreciated that this is an extremely close balance, the maintenance of which, under the different conditions of temperature and humidity to which the bridge may be subjected, requires careful consideration of the effects of these factors.

The effective phase-angle balance, though discussed above in terms of capacitance only, is, of course, the resultant of the inherent residual inductances and capacitances of the coil windings plus the additional capacitance effects due to the coil shields. The component due to residual magnetic induction is not appreciably affected by temperature or frequency changes. The capacitance component of the winding, however, tends to vary with temperature in accordance with the temperature coefficient of capacitance of the dielectric used for insulating the wire and with frequency to the extent that the capacitance is affected by absorption.

It is common practise to employ silk-insulated wire treated with varnish or wax for purposes of protection against moisture in such coils. In order to obtain data covering the temperature and absorption effects and also the phase-angle characteristics of silk insulation, both untreated and when treated with a number of the more common materials, various samples were constructed and tested as indicated in Table I.

TABLE I

CAPACITANCE AND PHASE-ANGLE TESTS BETWEEN TWO DOUBLE-SILK-COVERED NO. 38 A. W. G. WIRES, EACH 88 IN. (224 CM.) LONG, WOUND IN BIFILAR FASHION ON A GLASS TUBE ONE IN. (2.54 CM.) IN DIAMETER AND THEN TREATED WITH VARIOUS MATERIALS. SAMPLES DRIED BEFORE TESTING.

Test sample number	Material used for treatment	Capacitance, micro-microfarads				Phase-angle tangent			
		Temp.—20 deg. cent.		Temp.—45 deg. cent.		Temp.—20 deg. cent.		Temp.—45 deg. cent.	
		1000 ∾	50,000 ∾	1000 ∾	50,000 ∾	1000 ∾	50,000 ∾	1000 ∾	50,000 ∾
1	None	193.3	188.2	188.7	184.1	0.0083	0.013	0.0076	0.011
2	Paraffin	251.1	243.2	232.1	244.3	0.0095	0.014	0.010	0.011
3	Collodion	238.9	228.0	231.4	222.5	0.016	0.021	0.015	0.018
4	Beeswax compound	258.2	248.7	273.2	265.8	0.013	0.014	0.010	0.013
5	Pyralin	253.1	241.1	258.5	246.6	0.016	0.021	0.018	0.022
6	Insulating varnish	346.4	320.1	361.5	337.0	0.027	0.036	0.030	0.033
7	Shellac	296.9	284.0	314.3	302.4	0.012	0.021	0.015	0.020

From the standpoint of percentage capacitance change (reckoning from the minimum temperature and frequency conditions as being those at which initial adjustments would be made), untreated and paraffin-treated silk insulation were found to be appreciably superior to any of the other materials. The change due to absorption effect (about three per cent for these two) was considered the more important, as normally variations in temperature would not be very large. As would be expected, the untreated silk had the lowest phase-angle effect and also the smallest capacitance. Assuming that a method of excluding moisture could be devised, it was concluded that an untreated silk-insulated winding would be the best to use, although the paraffin treatment was also considered promising. Discounting the fact that the two ratio coils would change in the same direction though not necessarily by the same amount, it was decided that a satisfactory factor of safety would be provided if it were assumed that the coils might become unbalanced by one half the observed change in one coil; that is, by about 1½ per cent. In order that such unbalance should not exceed 0.1 mmf., the capacitance of each coil would need to be not more than about 6.0 mmf. It should be noted that this limit applies to the true inherent capacitance and not to the resultant of the coil capacitance and inductance.

Considering now the variation in resistance over the frequency range of the bridge, it can be shown, following the methods of Curtis and Grover, that the effect of a capacitance of the value noted

above on the resistance will not exceed one part in 100,000, which is quite satisfactory. The change in resistance (from the d-c. value) due to energy dissipation in the insulation is, however, somewhat larger than that due to the pure capacitance effect. This change is given to a close approximation by the expression

$$\Delta R = \frac{C\omega R^2 \tan\phi}{3},$$

where C is the total distributed capacitance between the wires of a bifilar winding and ϕ is the phase angle of the capacitance.[7] Clearly both the capacitance and its phase angle should be kept as small as practicable. An obvious and simple way of attaining the first object would be by using the very finest wire available. To do this, however, would in many cases result in excessive heating of the resistance. For bridge tests on telephone apparatus the ratio arm current will rarely exceed 25 milliamperes. The energy to be dissipated in a 1000-ohm coil is then about 0.5 watt, requiring a radiating surface of about 25 sq. cm. for a maximum temperature rise of 10 deg., which is a desirable limit. Since only the outer surface of such a coil is effective in radiating the generated heat the question of the number of layers requires consideration. Other factors being constant, it has been found that of the various possible arrangements that are easily constructed and mounted, a sectionalized, two-layer winding gives minimum capacitance. Hence one half of the winding is required to have an exposed surface of 25 sq. cm. The gauge of wire is then determined as a function of its specific resistance. A resistance alloy having a suitably low temperature coefficient (such as manganin, advance, etc.) will, on this basis, require that a wire no smaller than No. 38 A. W. G. be used. This is the size of wire used by Curtis and Grover and in the experiment covered by Table I. Using the data of this table, it was calculated that a Curtis and Grover type of coil, except for treatment, would have a change in resistance of not over one part in 50,000. The lower capacitance coils required from the phase-angle standpoint would have even smaller changes.

Summing up, then, the ratio arm coils were to be of approximately 1000-ohm resistance, wound with No. 38 A. W. G. double-silk-insulated manganin, or advance resistance wire, dried, but not impregnated with any moisture-resisting compound; the winding was to be arranged so that the true capacitances would not exceed 6.0 mmf. Besides being balanced for d-c. resistance, the resultants of their capacitance and inductance values were to be balanced to within 0.1 mmf. To

[7] See Note 6.

meet these requirements, the bridge coils were constructed as follows. The spool used is a glass cylinder $\frac{3}{4}$ in. in diameter. The winding is applied as follows: Starting at one end of the spool, a single strand of the wire is wound on until 14 inductive turns have been applied giving a resistance of approximately 50 ohms. Then the wire is tied, the direction of winding reversed, and an exactly equal number of turns wound over the first 14, but in an opposite direction. This brings the wire to the beginning. It is again tied, carried parallel to the axis of the spool over this first section and a second section wound. This is continued until ten sections have been applied. A thin sheet of mica is tied in place around the winding and the projecting ends of the wire bared of insulation. The whole is then baked to anneal the wire and dry the insulation. While hot, it is dipped several times in molten asphalt compound until a continuous coating of this moisture-proof material has been formed over the winding and surrounding mica wrapping. Adjustment for resistance balance is made by varying the length of the two wire ends.

The effective reactances of coils made as above are positive before assembly in their shields. The effect of the shield is to increase the capacitance. Table II gives data obtained on the two coils made for

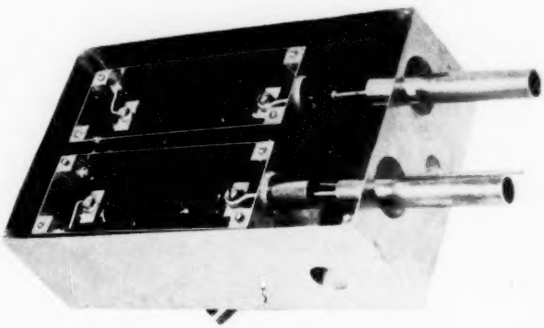


Fig. 12—Ratio arms

the bridge and shows the uniformity of phase-angle difference maintained by these coils over the operating frequency range. Final adjustment for reactance balance is made with the coils in the bridge circuit, a small amount of inductive coiling of the terminal leads sufficing for this purpose. In establishing this balance, use is made of the reversing switch described in the following section. Fig. 12 shows these coils assembled in their shields.

TABLE II

EFFECTIVE INDUCTANCE OF RATIO ARM COILS

Test Frequency	Microhenrys					
	Before Coating		After Coating		Assembled in Shield	
	Coil *A*	Coil *B*	Coil *A*	Coil *B*	Coil *A*	Coil *B*
1,000 cycles	7.4	6.9	6.7	6.3	− 1.3	− 1.0
50,000 cycles	7.4	6.9	6.7	6.2	− 1.2	− 0.8

Resistance of each coil = 1051.2 ohms

Reversing Switches. As will be noted from the diagram showing the circuit arrangement of the reversing switches, these are required to be completely enclosed in a shield which is connected to the junction point D of the bridge. They must also, of course, be subject to manipulation.

Obviously, then, this shield must be supported in some fashion from the outer enclosing ground shield of the bridge. The admittance between these two shields is a direct shunt on either one half or all of the impedance arm CD. While the capacitance component of this admittance can be readily balanced, it is more difficult to balance the conductance component, since the latter varies irregularly with frequency. Consequently, it is desired to make this factor so low that it can ordinarily be neglected. The construction adopted for this purpose is shown in the illustration, Fig. 13, which is a partially assembled view of the two reversing switches, their shield and its supporting brackets. It will be noted that the shield is supported by the brackets by means of small glass rods (four of which are shown). The low phase-angle characteristics of glass make it a favorable material to use, from this standpoint, but from the standpoint of machining into shape suitable for insulating supports, it is not so good. The construction shown, however, adapts it very well to this purpose. One other feature is worthy of special note; that is, the small change in position of any of the switch parts which occurs in effecting a reversal, the only metallic part moving being the small metal segments of the rotating disk.

Transformers. The transformers used for isolating the bridge circuit electrostatically from the source of testing current and from the detector system should have substantially zero external electromagnetic fields. This is to prevent inductive coupling to other parts

of the bridge circuit. For this purpose the transformer core is made in toroidal or ring form and the windings, both primary and secondary, are uniformly distributed about its circumference. The wound toroid is also completely enclosed in a sheet iron case.

The winding which is connected to the bridge has an electrostatic shield completely surrounding it for the purpose of concentrating all

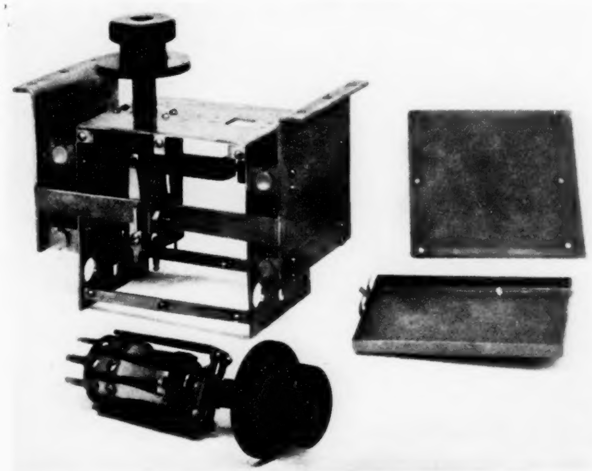


Fig. 13—Reversing switches

capacitance currents at one point. Around this localizing shield there is a second or ground shield. These two shields are made of sheet copper approximately No. 30 gauge (0.010 in.) in thickness. The inner winding terminal leads are brought out through a small brass tube leading into a terminal chamber which is an extension of the localizing shield. Since the admittance between the localizing and ground shields forms a major part of one of the balanced admittances shunting the impedance arms, it is desirable that the capacitance component be of low value and essential that it be constant. The conductance component should be negligibly small. To attain this end, the shields are separated at definite distances by means of hard rubber rings turned to fit the outer corners of the inner shield and the corresponding inner corners of the enclosing shield. These rings are made of the smallest cross-section consistent with mechanical strength requirements so as to introduce the minimum amount of solid material into the space between the shields. This minimizes the capacitance

and conductance values. These shields must not, of course, be allowed to act as short-circuited secondaries on the transformer which would be the case if they linked conductively with the windings. Each is therefore made in two parts similar to toroidal channels which upon assembly have their overlapping inner circumferences insulated from each other by means of thin mica laminations. Further details of the construction will be evident from a study of Fig. 14.

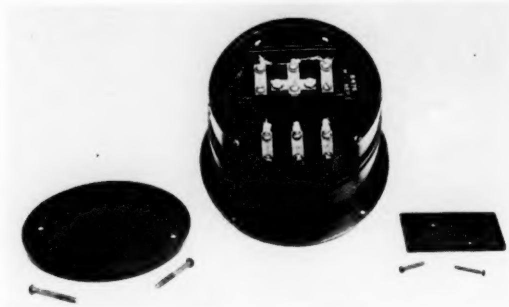


Fig. 14—Shielded transformer

The windings are, of course, proportioned so as to connect with a reasonable degree of efficiency the associated impedances. For best results two sets of transformers are used to cover the complete frequency range, one from 500 to 5000 cycles and the other from 5000 to 50,000 cycles.

Balancing Condensers and Impedance Arm Balance. It has been brought out previously that two adjustable capacitances are required, one to effect the proper adjustment of the bridge capacitances to ground and the other to balance the residual capacitances shunting one of the impedance arms. Such capacitances are provided in the form of adjustable air condensers each having a maximum value of about 500 mmf. The construction used is such as to give a high degree of stability of capacitance combined with low conductance characteristics. The arrangement of these condensers in relation to the other apparatus is shown in Fig. 3.

The effect on the accuracy of the bridge of the degree of balance of the impedance arms obtained by means of the balancing condenser C_b is determined as follows:

Capacitance Shunting the Impedance Arms. The equations giving the equivalent series inductance L' and resistance R' of a reactance of

inductance L and resistance R paralleled by a capacitance C are

$$L' = \frac{L - CR^2 - \omega^2 CL^2}{(1 - \omega^2 CL)^2 + \omega^2 C^2 R^2},$$

$$R' = \frac{R}{(1 - \omega^2 CL)^2 + \omega^2 C^2 R^2}.$$

When the bridge is balanced the equivalent series values of each component of the two impedance arms must be equal respectively to each other. If, however, the two arms have different shunting capacitances, it is evident that this equality will be obtained only by making the values of the two inductive branches of the parallel circuit somewhat different from each other. This difference represents the error introduced by the capacitance unbalance. When the values of the shunting capacitances are small these errors for the purpose of indicating their order are sufficiently closely given by the expressions

$$\Delta L_X = \omega^2 L_X^2 (C_X - C_S) \tag{7}$$

and

$$\Delta R_X = 2\omega^2 L_X R_X (C_X - C_S) \tag{8}$$

where C_X and C_S are the capacitances shunting the unknown and standard impedance arms, respectively. Reduced to percentages, these expressions become

$$\Delta L_X\ (\%) = 100\omega^2 L_X (C_X - C_S)$$

and

$$\Delta R_X\ (\%) = 200\omega^2 L_X (C_X - C_S)$$

and may also be written

$$\Delta L_X\ (\%) = 100\frac{C_X - C_S}{C_R} \tag{9}$$

and

$$\Delta R_X\ (\%) = 200\frac{C_X - C_S}{C_R}, \tag{10}$$

where C_R is the value of capacitance that would be required for resonance with inductance L_X at the test frequency.

These errors are thus proportional to the ratio of the capacitance unbalance to the resonating capacitance of the inductance under test. Ordinarily, values of the latter factor do not go below about 500 mmf. so that in the worst case a difference in capacitance of 0.1 mmf. corresponds to errors of 0.02 per cent and 0.04 per cent in inductance and resistance respectively.

Shields and Wiring. The shields have sufficient rigidity and are supported so as to maintain a definite and constant space relation to the part shielded and to the other shields. They are also of sufficiently high conductivity to maintain a common definite electrical potential at all points with respect to the part shielded.

The supports of shields or of bridge elements within the shields are as nearly as possible of constant specific inductive capacity, have low dissipative and leakage losses and are restricted to the minimum in number and size consistent with meeting the required rigidity of support.

Interconnecting conductors are shielded within brass tubes of approximately ½-in. diameter, the conductor which is of No. 10 gauge copper being supported at the axis of the tube by means of glass beads fitting snugly within the tubes and having holes through which the conductor passes. These beads are located longitudinally on the conductor by means of a small lump of solder placed on each side.

Standards. The impedance standards consist of adjustable self-inductance elements used in series with an adjustable non-inductive resistance. Each self-inductance element consists of a series of inductance coils and a low range inductometer of the Brooks type,[8] arranged in three decade formation and connected to dial switches by means of which any series combination of the coils can be selected. The inductometer is always in circuit and permits of balancing inductance values that fall between consecutive steps on the dials. Fig. 15 shows, schematically, the connections used for these standards and also the way in which they are shielded. It will be noted that the parts comprising each decade have a shield enclosing them and also all preceding decades of higher value. This makes a rather complicated mechanical arrangement but results in very important advantages from the standpoint of electrical performance. Due to the individual decade shields, each decade has effective values that are entirely independent of the settings of either of the other decades. Hence, once each individual setting of each dial has been calibrated, the value for the standard as a whole for any possible combination is obtained by simple addition of the separate dial values. This saves an immense amount of work in calibrating and also simplifies the reading of the standard. Without these shields the inter-coil and coil-to-ground admittances, at the higher frequencies, are sufficiently large to make the effective impedance of each decade setting depend to an appreciable extent upon the settings of the other decades. Under such conditions a calibration of

[8] H. B. Brooks and F. C. Weaver: "A Variable Self and Mutual Inductor," *Scientific Paper of the Bureau of Standards*, No. 290.

every combination would be required, and as this calibration would vary with frequency, a correction would be needed for each frequency value used.

In a standard of inductance to be used at high frequencies it is, of course, always desirable in order to minimize capacitance effects to have the inductance coils as small as possible. In the case of a completely shielded decade standard this is even more important on

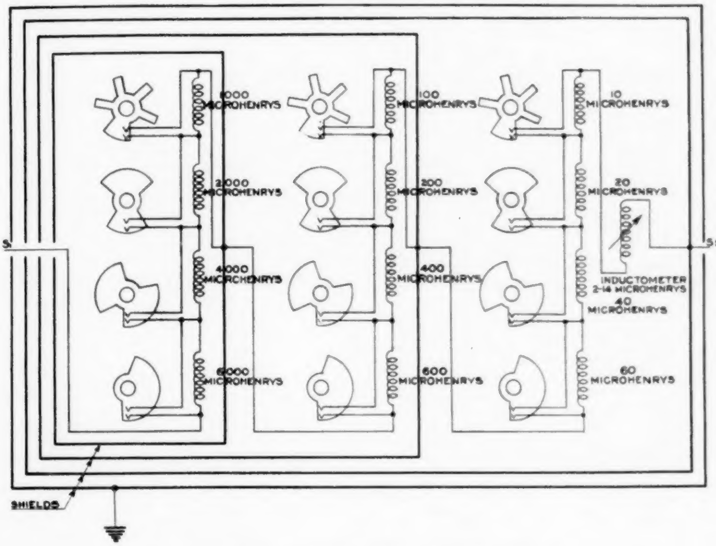


Fig. 15—Circuit diagram of shielded inductance standard

account of the capacitances added by the internesting shields. On the other hand the coil resistances should be quite small in comparison with their reactances, a requirement which tends to increase the coil dimensions. In addition to the above such coils should be highly stable in their inductance and effective resistance values with respect to the residual effects of direct and alternating currents and of temperature and humidity changes. Their values should also be of a satisfactory degree of constancy with respect to frequency and value of the testing current.

To meet these varied requirements the coils used in this bridge depart from the air core type ordinarily employed, in that they have a magnetic core of high stability and efficiency. Thus the desired inductance is obtained with a much smaller number of turns in the

winding giving a satisfactorily low resistance even in a coil only a fraction of the size of the equivalent air core coil. An adaptation of the new magnetic material, permalloy, has made this type of inductance standard possible.[9] Their cores consist of finely laminated, high

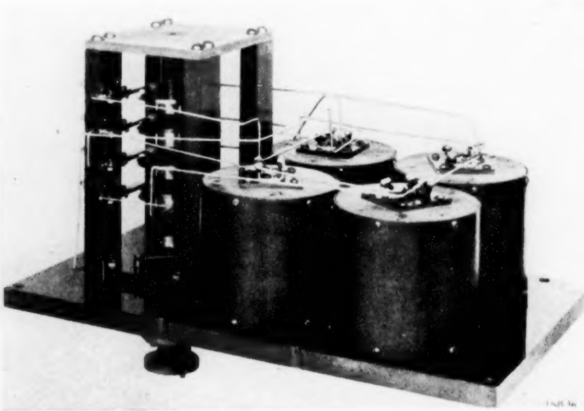


Fig. 16—Coil and dial switch assembly of typical inductance standard decade

specific resistance permalloy punchings, carefully annealed and assembled to form a toroidal structure whose effective permeability is about forty. On this is wound a sectionalized winding of insulated stranded conductor, the individual strands also being insulated from each other. The wound coils, after adjustment to the value desired, are sealed with moisture-proof compounds in phenol fiber cases. Fig. 16 shows an assembly of the four coils and switch which comprise one decade of the standard. In Table III are given data for typical coils illustrating their performance in respect to the above points.

The adjustable, non-inductive resistance is a commercial dial resistance box to which a shield has been added. It has five dials providing a range of 1000 ohms in steps of 0.01 ohm. Its shield is grounded in use, the resistance itself being connected usually between the *C* corner of the bridge and the inductance standard but in the case of an unknown impedance having a lower resistance than the standard from the *C* corner to the coil under test.

[9] H. D. Arnold and G. W. Elmen, *Franklin Institute Journal*, 195, 1923.

TABLE III

DATA ON COILS FOR INDUCTANCE STANDARDS

	Frequency Range	
	500 ∾ — 5,000 ∾	5,000 ∾ — 50,000 ∾
Overall Dimensions		
Diameter of case........	6½ in.	3½ in.
Length of case..........	4 in.	3⅛ in.
Inductance Characteristics		
Nominal value..........	0.100 henry	0.010 henry
Change with frequency..	+0.5% (500 ∾ — 5,000 ∾)	+2.5% (5,000 ∾ — 50,000 ∾)
Change with current....	+0.01% per milliampere	+0.007% per milliampere
Temperature coefficient..	−0.013% per deg. Fahr.	−0.005% per deg. Fahr.
Residual magnetization effect of one ampere d-c................	Less than 0.01%	Less than 0.01%
Resistance Characteristics		
At an alternating current (effective value) of		
2.0 milliamperes...	1,000 ∾ — 5.2 ohms 3,000 ∾ — 8.5 " 5,000 ∾ — 13.0 "	10,000 ∾ — 3.6 ohms 30,000 ∾ — 13.6 " 50,000 ∾ — 34.0 "
10.0 milliamperes...	1,000 ∾ — 5.5 " 3,000 ∾ — 9.1 " 5,000 ∾ — 13.9 "	10,000 ∾ — 3.7 " 30,000 ∾ — 13.9 " 50,000 ∾ — 34.7 "
Temperature coefficient............	−0.017% per deg. Fahr. at 3,000 ∾	− < 0.01% per deg. Fahr. at 30,000 ∾

PERFORMANCE

As was stated earlier, the operation of the bridge involves an initial balancing of its capacitances. It is then ready for impedance testing which is done by suitably connecting the unknown and standard impedances to the proper terminals and adjusting the latter until a balance of the bridge is obtained. The corresponding constants of the two impedance arms are then taken as being equal. Those of the standards being known, by calibration, it follows that those of the impedance under test can be simply derived. The degree of precision obtained depends upon two major factors, the accuracy of the calibration of the standards and the accuracy of the bridge comparison. The matter of calibration is beyond the scope of this paper and it will be assumed that a suitable calibration of the standards is available.

Due to the construction used the factors determining the accuracy of the bridge comparison of impedances are reduced to the following:

1. The resistance balance of arms AB and AD.
2. The effective shunt capacitance balance of these arms.

3. The direct capacitance balance of arms BC and CD.
4. The direct conductance balance of arms BC and CD.
5. The series inductance balance of the interior wiring to the impedance terminals of arms BC and CD.

As was explained in the foregoing two switches are provided for independently reversing the ratio arms (AB and AD) and also the outside connected impedances. These, therefore, afford a very convenient means of checking the above balances of the bridge network. By a suitable choice of the test condition under which the reversals are made, a fairly good approximation of the effect of the separate items can be made. The following series of tests indicate how this was done on one of these bridges.

The junction point C was first grounded. Then, with a telephone receiver as the detector and with a test current having a frequency of 1600 cycles, the setting of the condenser C_b was varied until a balance was obtained. The arms Z_S and Z_X were both open-circuited in this test; hence the capacitances shunting these arms alone determined the balance point. This balance was very sharp indicating that the shunting conductances were either very small or else accidentally well balanced. Leaving the condenser set at its balance point, there was then connected into one of the impedance arms a toroidal self inductance standard having a nominal inductance of 0.200 henry and an effective resistance of about 50 ohms. In the other arm there was connected a similar standard of the same nominal but of slightly lower actual value in series with a small adjustable inductance and adjustable resistance, each of sufficient range to effect a balance of the corresponding constants. The extension inductance was graduated in steps of one microhenry and the resistance in steps of 0.001 ohm. Balances for the four combination settings of the reversing switches, S_R and S_Z, were then made, only the extension elements being varied in getting these balances. Readings as given in Table IV were obtained.

TABLE IV

Switch Position		Extension Inductance	Extension Resistance
S_R	S_Z		
Right	Right	124 ± 2 microhenrys	4.00 ± 0.01 ohm
Left	Right	120 " " "	4.00 " " "
Right	Left	124 " " "	3.87 " " "
Left	Left	128 " " "	3.87 " " "

Consideration of these figures led to the following conclusions:

1. The change in inductance balance due to reversal of the ratio arms is not more than eight parts in 200,000 or 0.004 per cent. It has been shown previously that the phase-angle balance of the ratio arms is not critical with respect to inductance readings. Hence, the change in inductance may be considered to be closely indicative of the resistance unbalance of the ratio arms. From the above data it is seen that this does not exceed 0.01 per cent.

2. The change in resistance due to reversal of the ratio arms being within the limits of observational error, the phase angles of the coils themselves are, as nearly as can be determined by this test, exactly balanced.

3. Since the change of inductance balance due to reversal of the impedance arms is no more than that due to the ratio arm reversal, the capacitance balance of the impedance arms is apparently satisfactory. It should be noted, however, that this balance is not critical under these test conditions. (See eq. 9.)

4. Since the resistance balance was appreciably affected by reversal of the impedance arms, it appeared that there was an unbalancing factor present which, if affecting the ratio arms, was not reversed with them but which if present in the impedance arms was reversed. The latter might have been an unbalance of the impedance arm shunt conductances but it was assumed that this unbalance was quite small. On the other hand, as was mentioned earlier in the paper, there are two small inter-shield capacitances shunting the ratio arms which are not reversed by the ratio arm switch and it seemed likely that an unbalance of these capacitances was causing the change in resistance reading. This proved to be the case, as adjustment of the balance of these capacitances for which, as previously noted, provision had been made, resulted in identical resistance readings being obtained for both positions of the impedance arm switch. After this adjustment had been made the previous tests were repeated, resulting in readings as given in Table V.

TABLE V

Switch Position		Extension Inductance	Extension Resistance
S_R	S_Z		
Right	Right	124 ± 2 microhenrys	3.93 ± 0.01 ohm
Left	Right	120 " " "	3.93 " " "
Right	Left	124 " " "	3.93 " " "
Left	Left	126 " " "	3.93 " " "

As a further check on the performance of this unit, two inductances, each of about 0.01 henry inductance, were compared at two frequencies, 25,000 and 50,000 cycles. Table VI gives the readings obtained in these tests.

TABLE VI

Frequency Cycles	Switch Position		Extension Inductance	Extension Resistance
	S_R	S_Z		
25,000	Right	Right	123 ± 1 microhenry	5.1 ± 0.1 ohm
	Left	Right	122 " " "	5.1 " " "
	Right	Left	129 " " "	5.1 " " "
	Left	Left	128 " " "	5.1 " " "
50,000	Right	Right	38 " " "	24.2 " " "
	Left	Right	36 " " "	24.2 " " "
	Right	Left	57 " " "	23.8 " " "
	Left	Left	57 " " "	23.8 " " "

At the 25,000-cycle frequency the maximum difference in inductance from the probable correct balance does not exceed ± 5 microhenrys or 0.05 per cent. The resistance balances check to within 0.1 ohm. At 50,000 cycles, the inductance change due to ratio arm reversal is still within ± 0.01 per cent while the resistance change is within 0.1 ohm which would be just under one per cent for a coil of this reactance and a reactance to resistance ratio of 300. This is a critical test of the ratio arm phase-angle balance. Hence it may be concluded that over the entire frequency range the ratio coils meet all balance requirements. The changes in inductance occurring at the higher frequencies when the impedance arms were reversed indicated that the residual capacitance unbalance of these arms was too large. Readjustment of the balancing condenser reduced the changes to less than 0.02 per cent. The difference in resistance balance at the 50,000-cycle frequency indicates that the conductances shunting the impedance arms are not negligible at this frequency. For more accurate results these conductances would require balancing. This would be quite practicable by means of a variable high resistance shunt.

In making each series of tests outlined above, the testing potential applied to the bridge was varied by means of a resistance potentiometer from the lowest value at which a balance could be made to the maximum of the supply oscillator. This was to check the completeness of the shielding and to detect the presence of any coupling with the supply circuit. In no case was there any discernible change in balance produced.

Conclusion

A system of electrostatic shielding for a direct reading bridge for the measurement of inductive impedances at frequencies up to 50,000 cycles has been described.

The general considerations defining the balances of the various capacitances which this shielding controls have been discussed and specific requirements derived for a typical range of impedances. The physical construction of a bridge designed to meet these requirements has been described and test data given illustrating its performance. These have shown it to be capable of comparing impedances over the above frequency range with a precision which approximates that ordinarily found in routine direct current resistance measurements.

Letters to the Editor

From Mr. Arne Fisher: A Relation Between Two Coefficients in the Gram Expansion of a Function

From Dr. W. A. Shewhart: A Reply

From Mr. Fisher: A Further Note

To the Editor of the Bell System Technical Journal:

In a number of valuable and interesting contributions to this *Journal*, Dr. W. A. Shewhart has made an extended use of the infinite series of Gram. With all the controversy that at present is going on between the pure empiricists, attempting on the one hand to dragoon statistical analysis into a mere *inductio per simplicem enumerationem*, and the a priori theorists on the other hand, who claim that statistical methods so-called are nothing more than simple and evident applications of well-known principles of the probability calculus as formulated by Laplace, it has been a source of satisfaction to me to note that Dr. Shewhart apparently has given the latter methods a place of preference over the methods of the out and out empiricists.

Because of the fact that I happen to be responsible for having called the attention of English-speaking readers to the series of Gram and to have emphasized that Gram's development anteceded the less general developments by Edgeworth and the very special formula by Bowley by more than 20 years, I hope that I may be afforded an opportunity through the medium of your *Journal* to point out in brief form a few decidedly simple features of the Gram series which greatly add to its practical applications in statistical work.

Moreover, it seems that Dr. Shewhart, as well as other students in this country, have received a somewhat different idea about the nature of the Gram series than that which it was my intention to convey in my book on "The Mathematical Theory of Probabilities." This probably is my own fault. For while I have given in the above-mentioned book a description of the various methods for determining the coefficients of the individual terms of the Gram series, I did not mention the various degrees of approximations according to the number of terms as retained in the series itself. The reason for this omission is due primarily to the fact that I expect to treat this aspect in a forthcoming second volume of the book on probability in connection with the presumptive error laws of the a posteriori determined semi-invariants, which laws contain as a special case the evaluation of the standard (or probable) errors of the constants of the frequency curves.

The omission on my part to properly emphasize the close relation between the theory of sampling (i.e., the a posteriori probability theory) and the Gram series is probably also responsible for the fact that Dr. Shewhart in several of his articles has intimated that two terms in the Gram series in certain instances yield a better approximation than three or more terms. This idea has probably arisen from the mistaken notion on the part of Bowley of the generalized probability curve, which is a special example of the general Gram series. The following brief remarks should, therefore, not be taken as a criticism of Dr. Shewhart's work, but rather as a sort of amplification of some of the chapters in my own book on "The Mathematical Theory of Probabilities."

Gram's series, like the Fourier series, offers a perfectly general method for the expansion of arbitrary functions and is, contrary to the opinion of some students, not limited to frequency functions, although it there happens to be especially useful.

The underlying principles of the Gram series may be set forth briefly as follows: Let $F(x)$ be the true (or presumptive) function, which is known from either purely *a priori* considerations, or from observations, and let $G(x)$ be another function (the so-called generating function), which gives a rough approach to $F(x)$. Then according to Gram's method, we have

$$F(x) = c_0G(x) + c_1G'(x) + c_2G''(x) + \dots + c_nG^n(x). \qquad (1)$$

The generating function $G(x)$ may assume a variety of forms. In the case of generalized frequency functions, it is customary to select as the generating function, $G(x)$, a quantity $z = h(x)$ which is normally distributed, and write $F(x)$ as [1]

$$F(x) = c_0\varphi_0(z) + c_1\varphi_1(z) + c_2\varphi_2(z) + \dots + c_n\varphi_n(z), \qquad (2)$$

where $\varphi_0(z) = \dfrac{1}{\sqrt{2\pi}}\, e^{-z^2:2}$ is the generator and $\varphi_1(z), \varphi_2(z) \dots \varphi_n(z)$ its derivatives.

When viewed from the theory of elementary errors as originally introduced by Laplace in his monumental work, "Theorie des Probabilities," the Gram series takes on special significance in the way in

[1] If $z = h(x) = (x - M):\sigma$, or a linear function of x, and if the origin of the co-ordinate system is laid at M with σ as its unit, we have the *special* case, or the Charlier A series of the well-known form

$$F(x) = N[\varphi_0(z) + \beta_3\varphi_3(z) + \beta_4\varphi_4(z) + \dots].$$

The various types of the frequency curves of Pearson may of course also be used as generators in the Gram series.

which the possible combinations of the "elementary errors" actually enter into the expansion. It can be shown that there exists a definite relationship between on the one hand the relative order of magnitude of the elementary errors and, on the other, the arrangement of the individual terms of the Gram series.[2]

This relationship was already established by Thiele. It was probably first concisely formulated by Edgeworth, and later on by Charlier and Jörgensen.

The various degrees of approximations can be expressed by the following schemata:

1st approximation $\varphi_0(z)$,
2d approximation $\varphi_0(z) + c_3\varphi_3(z)$,
3d approximation $\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + c_6\varphi_6(z)$,
4th approximation $\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + c_6\varphi_6(z) + c_5\varphi_5(z) + c_7\varphi_7(z) + c_9\varphi_9(z)$.

The first approximation is the usual normal curve. The second is the one which the English statistician, Bowley, erroneously thinks represents a generalized frequency function and for which Dr. Shewhart has shown a marked preference. The third approximation, except for the term involving the sixth derivative, has been used very extensively by Charlier.

Through the publication by C. V. L. Charlier in 1906 of extensive tables to four decimal places of the third and fourth derivatives, the Gram series was made available for practical statistical work in the case of frequency distributions with a moderate degree of skewness and excess (kurtosis). But although Charlier was aware of the fact that the retention of the fourth derivative—which is related to excess (kurtosis)—automatically brings about the inclusion of the sixth derivative, it was not before Jörgensen issued his large numerical tables of the first six derivatives to seven decimal places that we were able to do full justice to the third approximation of the Gram series. Incidentally it might in this connection be mentioned that it is doubtful if the much lauded test for "goodness of fit" as devised by Pearson

[2] Whenever we use the method of moments, the arrangement of the individual terms is *not* arbitraiy but must be made according to "order of magnitude" of the various derivatives; and the orders of magnitudes do not correspond to the indices of the derivatives. The generic term "order of magnitude" has in this instance only reference to the formation of the "elementary errors"; if taken in any other sense it is meaningless. The fourth and sixth derivatives are of the same order of magnitude; while the fifth, seventh and ninth all are of the next order following the fourth and sixth. The concept of the different orders of magnitude of the elementary errors is due to Poisson who already in 1832 arrived at the second approximation of the Gram series.

really is able to test the graduating ability of the Gram series as adequately as the more powerful, although far more complicated, "error critique" of Thiele. From Pearson's derivation it appears that his test is not able to take care of elementary errors beyond the first or second order, while it is necessary to consider the formation of elementary errors of the third order in the third approximation of the Gram series. In some work I have been doing in the way of construction of compound mortality curves, I have at least found that the Pearson test is inadequate, if actually not misleading, because it apparently fails to measure the effect of the elementary errors of higher order which enter into the formation of such compound mortality curves.

There exists, however, a very simple relationship between the coefficients c_3 and c_6 in the third approximation. We have, namely, with a fair approach to exactitude, the simple relation: $c_6 = \frac{1}{2}c_3^2$. It is therefore not necessary to calculate the semi-invariants or moments of higher orders than those of the fourth order, since we shall have

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + \tfrac{1}{2}c_3^2\varphi_6(z)$$

as a third approximation.

As an illustration of the above formula, we may select the expansion of the point binomial $(0.1 + 0.9)^{100}$. We have here, according to the formulas on pages 263–264 of my "Mathematical Theory of Probabilities":

$$s = 100, \qquad p = 0.1, \qquad q = 0.9$$

and

$$\lambda_1 = M = sp = 10, \qquad \sigma = \sqrt{spq} = 3, \qquad c_3 = -0.0444, \qquad c_4 = 0.0021$$

and

$$c_6 = \tfrac{1}{2}c_3^2 = 0.0010,$$

or

$$(0.1 + 0.9)^{100} = \tfrac{1}{3}[\varphi_0(z) - 0.0445\varphi_3(z) + 0.0021\varphi_4(z) + 0.0010\varphi_6(z)],$$

where

$$\varphi_0(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2:2}$$

and

$$z = (x - 10) : 3.$$

A comparison between the above approximation and the true expansion of the point binomial $(0.1 + 0.9)^{100}$ to 4 decimals is given in the following table.

x = No. of Successes	Gram Series	True Value	No. of Successes	Gram Series	True Value
0	.0000	.0000	13	.0744	.0743
1	.0003	.0003	14	.0515	.0513
2	.0016	.0016	15	.0327	.0327
3	.0060	.0059	16	.0192	.0193
4	.0160	.0159	17	.0105	.0106
5	.0338	.0339	18	.0054	.0054
6	.0594	.0596	19	.0026	.0026
7	.0888	.0889	20	.0012	.0012
8	.1149	.1148	21	.0005	.0005
9	.1305	.1304	22	.0002	.0002
10	.1318	.1319	23	.0001	.0001
11	.1198	.1199	24	.0000	.0000
12	.0988	.0988			

The approximation is in this case well nigh perfect and comes much closer to the true values of the point binomial than any of the six approximations as given in Dr. Shewhart's article in the January 1924 number of this *Journal*. It also shows that with exactly the same amount of computation as that involved in the so-called Charlier A series, we can reach greatly improved results through the inclusion of the sixth derivative in the series. This arises from the important fact that once we have computed the coefficients c_3 and c_4, it is not necessary to calculate c_6 since $c_6 = \frac{1}{2}c_3^2$ approximately. Moreover, since extensive tables, notably those of Jörgensen, now are available for the normal function and its first six derivatives, there seems no good reason why we should not use the more exact approximation than the inexact formula by Bowley.

In conclusion, it might be well to emphasize the fact that while it is important to consider the relative order of magnitudes of the separate terms in the Gram series when we use the methods of semi-invariants or of moments, such restrictions are not necessary if we use the method of least squares in conjunction with properly determined weights.

ARNE FISHER.

December 10, 1926.

To the Editor of the Bell System Technical Journal:

I have read Mr. Fisher's communication with considerable interest. We who do not read the Scandinavian language owe much to him for his very able amplification and interpretation of many important contributions of the Scandinavian school of mathematical statisticians and this debt has been increased by the above communication insofar as it brings to light a very interesting relationship (the discovery of

which is attributed to Thiele), namely, that in the notation of the communication the constant c_6 is approximately equal to $\frac{c_3^2}{2}$.

Mr. Fisher definitely states that no criticism of my work is intended, but incidental to bringing out the above relationship he makes certain statements upon which I should like to comment briefly.

He states that the omission on his part to properly emphasize a close relation between the theory of sampling and the Gram series is probably responsible for the fact that I have intimated that two terms of the Gram series in certain instances yield a better approximation than three or more terms. To my knowledge this is not the case.

The special form of the Gram series used in my published articles in this *Journal* is that represented by his Equation 2.[1] The validity of this expansion rests upon the Lebedeff theorem.[2] So far as I am aware I have not intimated that two terms of the series yield a better approximation than three or more terms in the sense that

$$|F(z) - [c_0\varphi_0(z) + c_3\varphi_3(z)]|$$

should be less than

$$|F(z) - [c_0\varphi_0(z) + c_3\varphi_3(z) + \dots + c_n\varphi_n(z)]|$$

irrespective of n, although it is in this sense that Mr. Fisher discusses his example of the graduation of $(.9 + .1)^{100}$. To have done so would have been an obvious blunder because, assuming the Lebedeff theorem to be true, the absolute value of the difference ϵ between the function $F(z)$ and the sum of the first n terms of the series can be made as small as we please by taking n sufficiently large.[3]

I did say, however, in my article in the October issue of this *Journal:* "Carrying out steps 1 and 2, we conclude that the best theoretical equation representing the data in Fig. 1 is either the Gram-Charlier series (2 terms) or the Pearson curve of Type IV for both of which the estimates of the parameters may be expressed in terms of the first four moments μ_1, μ_2, μ_3 and μ_4 of Fig. 3." Of course the first two terms of the Gram-Charlier series requires only μ_1, μ_2 and μ_3. "Best" as used here obviously is in the sense of probability of fit which is entirely different from saying that the first two terms is the best approximation in the sense discussed by Mr. Fisher at least as illustrated by his

[1] It is of course understood that, in practice, transformations are made so that c_1 and c_2 are both equal to zero. In what follows, therefore, the second term of the series will be $c_3\varphi_3(z)$.

[2] Fisher, Arne, "Mathematical Theory of Probabilities," 2d edition, 1922, p. 203.

[3] It can be seen from my published work, however, that the sum of two terms is sometimes better than the sum of three.

example. In this case I found that the probability of fit for two terms was greater than that for three. Now, I find that it is as good as for Mr. Fisher's third approximation. It may be of interest also to know that statistical distributions sometimes arise where the first three terms give as good a fit as Mr. Fisher's third approximation involving 4 terms. This is particularly true when the universe from which the sample is drawn is nearly symmetrical. My action in this connection can be justified both upon theoretical and practical grounds but we need not do more than mention this point to make sure that the reader will not confuse my statement quoted above with what Mr. Fisher is talking about in his communication.

Having thus dismissed the questions which may arise in connection with published work in this *Journal*, I should like to add a word or two of caution to the reader of Mr. Fisher's letter where it reads: "Moreover, since extensive tables, notably those of Jörgensen, now are available for the normal function and its first six derivatives, there seems no good reason why we should not use the more exact approximation than the inexact formula by Bowley."

We have made far more use of the Gram series in connection with our inspection work than indicated in the published papers. In this work we have found that it is theoretically not necessary in certain instances and in many more instances it is not practical to follow Mr. Fisher's suggestion. I shall limit my remarks to the application of the series which we have made in expanding a known function in terms of an infinite series in which the generating function is the normal law. In this connection the outstanding practical question is: Given the known function $F(x)$, what number n of terms of the infinite series must we take in order that the absolute magnitude of the difference between the function $F(x)$ and the sum of the n terms will be less than a given preassigned quantity ϵ? I am sorry that Mr. Fisher does not answer this question. Instead he proposes a grouping of terms upon the basis suggested in a footnote to his article. Now, it may easily be shown in the particular case cited by Mr. Fisher, i.e., the graduation of the point binomial $(.9+.1)^{100}$, that the sequence of signs depends upon the value of z, that for certain values of z his second approximation is just as good as his third, and that in many instances the difference between the second approximation and the third is not sufficiently great to be of any practical importance. Whether we should use the second, third, or higher approximation in a given case is one for special consideration.

In closing let me say that I have not made the above remarks with any intention of discrediting the applications of this series but rather

to indicate to the casual reader that there are certain technical questions involved in its application which must be given due consideration even beyond the stage outlined in Mr. Fisher's communication. I have found that this series often has many advantages over competing methods of analyzing data although not all of these advantages are referred to in the literature of the subject.

W. A. SHEWHART.

December 28, 1926.

To the Editor of the Bell System Technical Journal:

The question raised by Dr. Shewhart as to the measure of the absolute magnitude of the difference between a known function, $F(x)$, and the first n terms of the series has been treated by Gram in his original article on "*Rækkeudviklinger bestemte ved Hjælp af de mindste Kvadraters Metode.*" (On Development of Series by means of the Method of Least Squares.) In this article Gram also discusses at length the decidedly practical question of arriving at an estimate of the remainders (or residuary terms), which invariably occur in practice where we, of course, are forced to deal with a finite number of terms.

It would, however, be beyond the limits of the present communication to enter into this aspect of the question, which necessarily is somewhat complicated. In passing it, I wish merely to state that Gram's original method of determining the coefficients in the series on the basis of the principle of least squares is decidedly easier to apply than the relatively cumbersome method of moments in arriving at a reliable measure of the remainder of the series after, say, the n^{th} term.

Dr. Shewhart's further contention that two terms of the Gram series sometimes give as good a fit as three or even four terms, and that three terms in the case of nearly symmetrical distributions serves as well as four terms, seems to me to be almost self-evident from a simple consideration of the way in which the coefficients c actually enter into the series.

All the terms containing uneven indices tend to produce skewness, and all the terms with even indices produce excess (kurtosis). If the coefficient c_3 is not too large, and if c_4 is small as compared with c_3, it is evident that

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z)$$

will give about as good an approximation as

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + \tfrac{1}{2}c_3{}^2\varphi_6(z).$$

On the other hand, in nearly symmetrical distributions with a pro-

nounced excess (kurtosis), where c_4 is large as compared with c_3, it seems also reasonable that

$$F(x) = c_0\varphi_0(z) + c_4\varphi_4(z)$$

might in certain instances give as good a fit as

$$F(x) = c_0\varphi_0(z) + c_3\varphi_3(z) + c_4\varphi_4(z) + \tfrac{1}{2}c_3{}^2\varphi_6(z).$$

These aspects of the series have been discussed by Thiele.

ARNE FISHER.

January 10, 1927.

Abstracts of Recent Technical Papers from Bell System Sources

Loading for Telephone Cable Circuits.[1] D. W. Whitney. This paper summarizes the principal characteristics of the loaded telephone line and discusses the major improvements in loading. Up to 1900 there was a general avoidance of the use of cable in the toll telephone plant, due to the high attenuation and distortion of speech currents not experienced in open wire lines. By the use of the loading coil and the telephone repeater, a network of toll cables has grown very rapidly which now connects the large population centers of the Atlantic seaboard and the upper Mississippi Valley region.

Electric Annealing of Magnetic Materials for Telephone Apparatus.[2] W. A. Timm. This paper briefly describes the annealing equipment used by the Western Electric Company prior to 1909 and some years afterward at the Hawthorne plant, in contrast to the more recently installed electrical equipment.

The Problem of Secondary Metals in World Affairs.[3] F. W. Willard. Preliminary to the discussion of the secondary metal industry, the author gives a brief statistical survey of the rate of depletion of the world's primary metal resources along with a study of metal production in the United States.

The rapid rate of exhaustion of known resources is emphasized. Graphs also show the relation of prices and production of copper and lead in the United States over a period of 40 years.

Each important economic metal is discussed briefly showing how its present commercial uses affect its return through the secondary market. Attention is directed to the degradation of metals in use, forever eliminating them for re-use. Platinum, though not generally classified among economic metals, is treated briefly because of its key importance in certain industries. The seriousness of the present trend of using platinum in jewelry is indicated.

The growing importance of the secondary metal industry is shown graphically and briefly discussed, leading to an emphasis of the need

[1] To be published in the *Proceedings* of the Telephone and Telegraph Section of the American Railway Association.

[2] *Trans. American Soc. Steel Treat.*, p. 782, Nov. 1926.

[3] *Industrial and Eng. Chemistry*, p. 1178, Nov. 1926. Presented at the Round Table Conference on the rôle of chemistry in the world's affairs at Williamstown, Mass., Aug. 1926.

for encouragement of scientific research to increase metal recovery. The United States is favored above most nations in being largely self-contained with respect to original sources of economic metals, yet it has no source for tin and platinum and inadequate sources for manganese, chromium and other less common metals essential to present-day steels.

The work of a joint commission of the American Institute of Mining and Metallurgical Engineers and the Mining and Metallurgical Society of America is outlined, giving briefly their recommendations for international control of minerals.

The paper is concluded by emphasis of the importance of secondary metal recovery to national existence in times of stress when the nation may be thrown entirely upon its own resources. The conclusion also makes an appeal for the consideration of international conventions on economic metal exchanges and suggests that competent technical men be taken more into political councils in the treatment of a problem of this kind.

Tone Reproduction in the "Halftone" Photo-Engraving Process.[4] HERBERT E. IVES. The "halftone" photo-engraving process was invented, and its technique developed, prior to the days of what is commonly called "photographic sensitometry." The necessary conditions and the appropriate operations for securing "highlight" and "shadow detail" were found by empirical studies guided by the appearance of the result, as appraised by the unaided eye. No comprehensive sensitometric study of the halftone reproduction process appears to have been made, at any rate none have been published. The work described is a rough survey of the problem, but, due largely to the use of accurate photometric measurements, and the correlation of these measurements with other sensitometric data, rather decisive conclusions have been possible as to the essential characteristics of the process in question, and on the procedures necessary for its complete success.

Frequency Measurements with the Cathode Ray Oscillograph.[5] FREDERICK J. RASMUSSEN. The cathode ray oscillograph frequency measurement circuit described, differs from previous circuits in the use of by-pass condensers and plate leaks which permit the connection of the oscillograph to a.-c. circuits having large d.-c. components and which permit the use of biasing controls for shifting the position of patterns on the screen.

[4] *J. O. S. A. and R. S. I.*, March, 1926, p. 537.

[5] Presented before the *A. I. E. E.*, New York, N. Y., Nov. 1926.

Reference oscillators are used in conjunction with the frequency standards. They are of a type chosen for their high stability.

The well-known properties of Lissajous' figures are reviewed briefly and then are developed more fully for the cases in which only one term of their ratios may be determined from the oscillograph pattern. Following a general discussion of the accuracy of syntonization, there is discussed a detailed method of calibrating oscillators. The patterns used may be interpreted from one term of their ratio.

Several special circuits are described for use in frequency measurement work with the cathode ray oscillograph.

The methods and apparatus described are suitable not only for the technical measurements of a development and research nature but are equally adaptable for routine commercial work. The advantages which particularly commend themselves are the rapidity with which such work may be done and the ease with which the average man can learn the work.

Contributors to this Issue

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company, 1910–12; instructor of physics and electrical engineering, Princeton, 1912–14; American Telephone and Telegraph Company, Engineering Department, 1914–15; Patent Department, 1916–17; Engineering Department, 1918; Department of Development and Research, 1919–. Mr. Carson's work has been along theoretical lines and he has published several papers on theory of electric circuits and electric wave propagation.

John Davidson, Jr., B.E., Tulane University, 1906; E.E., Cornell University, 1909. Testing Department, General Electric Company, 1906–07; Construction Department, Allis Chalmers Company, 1907–1908; Electrical Department, Carnegie Steel Company, 1909–11; American Telephone and Telegraph Company Engineering Department, 1911–19, and the Department of Development and Research, 1919–. Since 1911 Mr. Davidson has been engaged in equipment development work and is now engaged on toll switchboard and signaling arrangements.

P. G. Edwards, B.E.E., Ohio State University, 1924; Western Union Telegraph Company, Traffic Department, 1919–21, Plant Department, 1921–22; American Telephone and Telegraph Company, Long Lines Plant Department, 1922–24; Department of Development and Research, 1924–. Mr. Edwards has been engaged in the development of toll test board equipment and toll test board methods.

H. W. Herrington, Washburn College, 1918–21; B.S. in E.E., University of Kansas, 1923; E.E., University of Kansas, 1926; American Telephone and Telegraph Company, Department of Development and Research, 1923–. Mr. Herrington has been engaged in the development of toll test board equipment and toll test board methods.

Karl K. Darrow, S.B., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., in physics and mathematics, University of Chicago, 1917; Engineering Department Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.

I. B. Crandall, A.B., Wisconsin, 1909; A.M., Princeton, 1910; Ph.D., 1916; Professor of Physics and Chemistry, Chekiang Provincial College, 1911–12; Engineering Department, Western Electric Company, 1913–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Crandall has published papers on infra-red spectroscopy, the condenser transmitter, thermophone, etc. More recently he has been associated with studies on the nature and analysis of speech which have been in progress in the Laboratory.

Lloyd Espenschied, Pratt Institute, 1909; United Wireless Telegraph Company as radio operator, summers, 1907–08; Telefunken Wireless Telegraph Company of America, assistant engineer, 1909–10; American Telephone and Telegraph Company, Engineering Department and Department of Development and Research, 1910–. Took part in long distance radio telephone experiments from Washington to Hawaii and Paris, 1915; since then his work has been connected with the development of radio and carrier systems.

William J. Shackelton, B.S. in E.E., University of Michigan, 1909; Western Electric Company, Manufacturing and Installation Department, 1909–10; Engineering Department, 1910–24; Bell Telephone Laboratories, 1925–. Mr. Shackelton's principal activities have been in connection with the design of loading coils and the development of methods of high frequency measurement.